

GenCore version 5.1.6 Copyright (c) 1993 - 2004 CompuGen Ltd.									
OM protein - protein search, using sw model									
Run on:	May 11, 2004, 13:23:11	; Search time 47 Seconds (without alignments) 120.233 Million cell updates/sec							
Title:	US-09-171-432A-38								
Perfect score:	109								
Sequence:	1 VDDPRSEEDKRFESHIECRK 20								
Scoring table:	BLOSUM62								
Gapop	10.0 , Gapext 0.5								
Searched:	1586107 seqs, 282547505 residues								
Total number of hits satisfying chosen parameters:	1586107								
Minimum DB seq length:	0								
Maximum DB seq length:	2000000000								
Post-processing:	Minimum Match 0% Maximum Match 100% Listing first 100 summaries								
Database :	A_Geneseq_29Jan04.* 1: geneseqp1980s.* 2: geneseqp1990s.* 3: geneseqp2000s.* 4: geneseqp2001s.* 5: geneseqp2002s.* 6: geneseqp2003as.* 7: geneseqp2003bs.* 8: geneseqp2004s.*								
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.									
SUMMARIES									
Result No.	Score	Query Match	Length	DB ID	Description				
1	109	100.0	20	AAW42921	Immunogen				
2	109	100.0	21	AAW42921	Synthetic				
3	109	100.0	854	AAW42921	Hepatitis				
4	109	100.0	1077	AAW42921	A partial				
5	109	100.0	1091	AAW42921	Translate				
6	109	100.0	2227	AAW42921	Attenuate				
7	109	100.0	2227	AAW42921	Hepatitis				
8	109	100.0	2227	AAW42921	Amino aci				
9	109	100.0	2227	AAW42921	Amino aci				
10	109	100.0	2227	AAW42921	Amino aci				
11	109	100.0	2227	AAW42921	Attenuate				
12	109	100.0	2227	AAW42921	Wild-type				
13	109	100.0	2227	AAW42921	Hepatitis				
14	109	100.0	2227	AAW42921	Attenuat				
15	109	100.0	2227	AAW42921	Attenuat				
16	109	100.0	2227	AAW42921	Wild type				
17	106	97.2	366	AAW42921	Sequence				
18	106	97.2	993	AAW42921	Sequence				
19	106	97.2	993	AAW42921	Sequence				
20	106	97.2	2227	AAW42921	Hepatitis				
21	106	97.2	2227	AAW42921	Hepatitis				
22	87	79.8	2227	AAW42921	Sequence				
23	81	74.3	839	AAW42921	Capsid re				
24	69	63.3	20	AAW42921	Immunogen				
25	69	63.3	21	AAW42921	Synthetic				

26	69	63.3	300	1	AAp90612	Portion 0
27	50	45.9	409	6	ABm69971	Photorhab
28	46.5	42.7	183	3	AAg57588	Arabidops
29	46.5	42.7	187	3	AAg57587	Arabidops
30	46.5	42.7	200	3	AAg57586	Arabidops
31	46	42.2	306	4	ABg22935	Novel hum
32	46	42.2	701	4	ABg01389	Novel hum
33	46	42.2	1663	7	ADb90024	Mouse com
34	44.5	40.8	1334	2	AAy02630	Protein e
35	44	40.4	131	3	AAg47694	Arabidops
36	44	40.4	134	3	AAg47693	Arabidops
37	44	40.4	231	2	AAy37694	Chlamydia
38	44	40.4	285	4	AAm78648	Human pro
39	44	40.4	292	4	AAm79632	Human pro
40	44	40.4	326	3	AAg50172	Arabidops
41	44	40.4	383	3	AAg50171	Arabidops
42	44	40.4	498	3	AAg50170	Arabidops
43	44	40.4	598	5	AAu72997	Neisseria
44	43.5	39.9	818	6	ADa33921	Acinetoba
45	43	39.4	125	3	AAg01681	Human sec
46	43	39.4	125	4	AAu23005	Novel hum
47	43	39.4	242	6	ABm70664	Photorhab
48	43	39.4	292	3	AAg26308	Arabidops
49	43	39.4	293	3	AAg26307	Arabidops
50	43	39.4	322	3	AAg26987	Arabidops
51	43	39.4	325	3	AAg50197	Arabidops
52	43	39.4	325	3	AAg32082	Arabidops
53	43	39.4	379	3	AAg26986	Arabidops
54	43	39.4	382	3	AAg50196	Arabidops
55	43	39.4	382	3	AAg32081	Arabidops
56	43	39.4	429	3	AAg42049	Arabidops
57	43	39.4	494	3	AAg26985	Arabidops
58	43	39.4	497	3	AAg50195	Arabidops
59	43	39.4	497	3	AAg32080	Arabidops
60	43	39.4	497	6	ABp81276	Arabidops
61	43	39.4	502	3	AAg42048	Arabidops
62	43	39.4	505	3	AAg42047	Arabidops
63	43	39.4	710	5	ABp69647	Human pol
64	43	39.4	780	4	AAw98947	Mitochond
65	43	39.4	780	4	AAb84333	Amino aci
66	43	39.4	780	7	ADd14186	Human src
67	43	39.4	802	4	ABg22298	Novel hum
68	43	39.4	814	4	ABg19229	Novel hum
69	43	39.4	1606	6	ABr41651	Human DIT
70	43	39.4	1715	4	ABb58089	Drosophil
71	43	39.4	2141	6	ABr41636	Human DIT
72	43	39.4	19938	6	ABp76682	Streptomy
73	42.5	39.0	405	7	ADe71290	Novel hum
74	42	38.5	20	2	AAw42919	Immunogen
75	42	38.5	21	4	AAb69436	Synthetic
76	42	38.5	133	6	ADb07114	Alloioctoc
77	42	38.5	145	4	ABb11556	Human bre
78	42	38.5	180	2	AAr22392	Antigen t
79	42	38.5	187	3	AAg26525	Arabidops
80	42	38.5	256	3	AAb42388	Human ORF
81	42	38.5	398	4	AAb95890	Human pro
82	42	38.5	453	4	AAm25915	Human pro
83	42	38.5	474	3	AAg27305	Arabidops
84	42	38.5	477	3	AAg27304	Arabidops
85	42	38.5	533	3	AAg27303	Arabidops
86	42	38.5	560	2	AAw18310	Human com
87	42	38.5	661	4	ABb61881	Drosophil
88	42	38.5	679	6	ABr52991	Protein s
89	42	38.5	704	4	ABb58127	Drosophil
90	42	38.5	704	6	ABu49771	Protein e
91	42	38.5	722	5	AAm47608	Drosophil
92	42	38.5	786	2	AAy55937	Human SUL
93	42	38.5	1001	2	AAy55942	Human/Mur
94	42	38.5	1001	5	ABb97326	Novel hum
95	42	38.5	1001	7	ADe55356	Human Pro
96	42	38.5	1005	6	ABr47509	Breast ca
97	42	38.5	1005	6	AAo30953	Human TAO
98	42	38.5	1166	7	ADb64795	Human pro

99 42 38.5 2759 6 AAO16418 Human nuc  
100 42 38.5 2781 3 AAY57453 Human tra

# ALIGNMENTS

RESULT 1  
AAW42921  
ID AAW42921 standard; peptide; 20 AA.  
AC AAW42921;  
XX  
DT 28-APR-1998 (first entry)  
XX  
DE Immunogenic Hepatitis A virus peptide YK-1314.  
XX  
KW Immunogenic peptide; immunogenic epitope; VP1 protein; immune response;  
KW antibody.  
XX  
OS Synthetic.  
OS Hepatitis A virus.  
XX  
FN WO9740147-A1.  
XX  
PD 30-OCT-1997.  
XX  
PF 18-APR-1997; 97WO-US006891.  
XX  
PR 19-APR-1996; 96US-0015644P.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI; 1997-535831/49.  
XX  
PT Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
PT response to HAV in a mammal or to detect the presence of antibodies  
PT against HAV in a mammal.  
XX  
PS Claim 13; Page 111; 140pp; English.  
XX  
CC Peptides AAW42906-21 are immunogenic peptides corresponding to  
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
CC substantially similar to a portion of the amino acid sequence of the VP1  
CC protein of HAV corresponding to amino acids 492-791. The present peptide  
CC is derived from amino acids 779-798, and has a reactivity of 54.2% with  
CC acute sera. Compositions containing the peptides can be used to induce an  
CC immune response to HAV in a mammal. The peptides can also be used to  
CC detect the presence of antibodies against HAV in mammalian serum. The  
CC peptides can also be used to make an antibody against HAV by  
CC administering the peptide to a mammal  
XX  
SQ Sequence 20 AA;

Query Match 100.0%; Score 109; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VDDPRSEEDKRFESHIECRK 20  
DB 1 VDDPRSEEDKRFESHIECRK 20

RESULT 2  
AAB69438  
ID AAB69438 standard; peptide; 21 AA.  
XX  
AC AAB69438;  
XX  
DT 20-APR-2001 (first entry)  
XX

DE Synthetic HAV VP1 peptide, SEQ ID NO: 38.  
XX  
KW Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
KW antigen; major structural capsid polypeptide; HAV antibody detection.  
XX  
OS Hepatitis A virus.  
OS Synthetic.  
XX  
FN WO200105824-A2.  
XX  
PD 25-JAN-2001.  
XX  
PF 14-JUL-2000; 2000WO-US019267.  
XX  
PR 15-JUL-1999; 99US-0144412P.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI; 2001-112681/12.  
XX  
PT Synthetic peptides used as antigen sources for enzyme immunoassays  
PT detecting anti-hepatitis A virus and as vaccines.  
XX  
PS Claim 10; Page 92; 130pp; English.  
XX

CC The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC Igm antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or biofluorescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy end  
CC of the peptides enhances the Igm antibody reactivity

XX Sequence 21 AA;

Query Match 100.0%; Score 109; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VDDPRSEEDKRFESHIECRK 20  
DB 1 VDDPRSEEDKRFESHIECRK 20

RESULT 3  
AAP50287  
ID AAP50287 standard; protein; 854 AA.  
XX  
AC AAP50287;  
XX  
DT 25-MAR-2003 (revised)  
DT 30-NOV-1991 (first entry)  
XX  
DE Hepatitis A virus (HAV) peptide corresponding to the capsid protein  
DE region of poliovirus RNA.  
XX  
KW Hepatitis A virus assay; antigen; antibody.  
XX  
OS Hepatitis A virus.  
XX  
FN WO8501517-A.  
XX

PD 11-APR-1985.  
XX  
PF 27-SEP-1984; 84WO-US001552.  
XX  
PR 30-SEP-1983; 83US-00537911.  
XX  
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX  
PI Ticehurst JR, Baltimore D, Feinstein SM, Purcell RH;  
PI Racaniello VR;  
XX  
DR WPI; 1985-098846/16.  
DR N-PSDB; AAN50330.  
XX  
PT New hepatitis A virus cDNA - useful in assays for the virus and for  
PT prodn. of the viral antigen and antibodies to it.  
XX  
PS Example; Fig 7; 60pp; English.  
XX  
CC The inventors claim HAV cDNA and a method for producing it, whereby large  
CC ams. can be obtd. economically. The cDNA is useful in the assay for  
CC detection of HAV quickly and easily and with high sensitivity and  
CC specificity. The HAV cDNA is also used in the prodn. of HAV antigen or  
CC antibodies to it. The antibodies may be monoclonal. (Updated on 25-MAR-  
CC 2003 to correct PA field.)  
XX  
XX Sequence 854 AA;  
SQ  
Query Match 100.0%; Score 109; DB 1; Length 854;  
Best Local Similarity 100.0%; Pred. No. 5.9e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 779 VDDPRSEEDKRFESHIECRK 798  
RESULT 4  
AAW95559  
ID AAW95559 standard; protein; 1077 AA.  
XX  
AC AAW95559;  
XX  
DT 28-APR-1999 (first entry)  
XX  
DE A partial hepatitis A virus (HAV) protein.  
XX  
KW Hepatitis A virus protein; HAV, P2 region;  
KW cell-culture-adapted HAV strain; infection; accelerated growth.  
XX  
OS Hepatitis A virus.  
XX  
PN US5849562-A.  
XX  
PD 15-DEC-1998.  
XX  
PF 06-JUN-1995; 95US-00468926.  
XX  
PR 30-SEP-1983; 83US-00537911.  
PR 27-SEP-1984; 84US-00654942.  
PR 06-OCT-1988; 88US-00256135.  
PR 06-NOV-1991; 91US-00788262.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Emerson SU, Purcell RH;  
XX  
DR WPI; 1999-094412/08.  
DR N-PSDB; AAX01006.  
XX  
PT Chimeric hepatitis A virus strains - with P2 region from cell-culture-  
PT adapted strain in wild-type genome.  
XX

PS Disclosure; Fig 7A-L; 36pp; English.  
XX  
CC The present sequence represents a partial hepatitis A virus (HAV)  
CC protein. The specification describes a DNA construct consisting of a wild  
CC type HAV genome in which the P2 region is replaced by the P2 region from  
CC a cell-culture-adapted HAV strain. The construct is used to demonstrate  
CC that mutations in the P2 region of a cell-culture-adapted HAV strain are  
CC sufficient for establishment of infection and accelerated growth in cell  
CC culture  
XX  
SQ Sequence 1077 AA;  
Query Match 100.0%; Score 109; DB 2; Length 1077;  
Best Local Similarity 100.0%; Pred. No. 7.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 1002 VDDPRSEEDKRFESHIECRK 1021  
RESULT 5  
AAR32426  
ID AAR32426 standard; protein; 1091 AA.  
XX  
AC AAR32426;  
XX  
DT 27-AUG-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 17-DEC-2001 (revised)  
DT 10-JUN-1993 (first entry)  
XX  
DE Translated from 5' region of Hepatitis A Virus genomic clone.  
XX  
KW HAV HM-175; chronic liver disease; picornavirus.  
XX  
OS Hepatitis A virus.  
XX  
FH Key Location/Qualifiers  
FT Region 1..711  
FT /note= "X's correspond to nonsense codons, i.e. this  
FT region is not an ORF"  
FT Region 238..1091  
FT /label= ORF  
FT /note= "second putative initiation codon at position 240"  
XX  
PN USN7788262-N.  
XX  
PD 15-DEC-1992.  
XX  
PF 06-NOV-1991; 91US-00788262.  
XX  
PR 30-SEP-1983; 88US-00536911.  
PR 27-SEP-1984; 84US-00654942.  
PR 06-OCT-1988; 88US-00256135.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
XX  
PI Ticehurst JR, Baltimore D, Feinstein SM, Purcell RH;  
PI Racaniello VR, Baroudy BM, Emerson SU;  
XX  
DR WPI; 1993-067429/08.  
DR N-PSDB; AAQ36934.  
XX  
PT Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn. of  
PT antigen and antibodies.  
XX  
PS Disclosure; Fig 7; 65pp; English.  
XX  
CC HAV virion RNA was extracted from the livers of marmosets which had been  
CC inoculated with HAV (the HAV had previously been passaged twice in  
CC marmosets). The RNA was used to prepare ds cDNA clones by standard  
CC methods. Clones contg. inserts which hybridised to RNA from HAV-infected  
CC

CC African Green Monkey kidney cells were selected for further analysis. A  
CC 7.4kb restriction map (about 99% of the HAV genome) was constructed from  
CC 5 overlapping inserts. The sequence of the first 3.3kb (approx.) from the  
CC 5'-terminus was determined. An amino acid sequence was deduced from the  
CC entire clone and an open reading frame was identified starting at  
CC position 238. A comparison of the predicted HAV amino acid sequences with  
CC the known capsid protein sequences of other picornaviruses (poliovirus,  
CC foot and mouth disease virus and encephalomyelitis virus) revealed areas  
CC of local homology. (Note: Revised entry submitted to correct the patent  
CC number format of US Government-owned NTIS applications to prevent clashes  
CC with ongoing US granted patent numbers. For further information please  
CC visit the Derwent web site at [www.derwent.com/dwpi/updates/ntis.us.html](http://www.derwent.com/dwpi/updates/ntis.us.html).)  
CC (Updated on 25-MAR-2003 to correct PI field.) (Updated on 27-AUG-2003 to  
CC correct OS field.)  
XX  
SQ Sequence 1091 AA;

Query Match 100.0%; Score 109; DB 2; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 7.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20  
|||  
Db 1016 VDDPRSEEDKRFESHIECRK 1035

RESULT 6  
AAR05697  
ID AAR05697 standard; protein; 2227 AA.

AC AAR05697;  
DT 24-OCT-2003 (revised)  
DT 27-AUG-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 15-AUG-1990 (first entry)  
XX  
DE Attenuated hepatitis A virus.  
XX  
KW Hepatitis A virus; vaccine; attenuated.  
XX  
OS Hepatitis A virus; strain HM-175.

Key	Location/Qualifiers
FT Region	1..23
FT Region	/label= VP4 = 1A
FT Region	24..245
FT Region	/label= VP2 = 1B
FT Region	246..491
FT Region	/label= VP3 = 1C
FT Region	492..791
FT Region	/label= VP1 = 1D
FT Region	792..980
FT Region	/label= 2A
FT Region	981..1087
FT Region	/label= 2B
FT Region	1088..1422
FT Region	/label= 2C
FT Region	1423..1496
FT Region	/label= 3A
FT Region	1497..1519
FT Region	/label= 3B = VPg
FT Region	1520..1738
FT Region	/label= 3C
FT Region	1739..2227
FT Region	/label= 3D

XX US4894228-A.  
XX  
XX  
PD 16-JAN-1990.  
XX  
XX  
PP 12-JUL-1988; 88US-00217824.  
XX

PR 19-SEP-1984; 84US-00652067.  
PR 09-SEP-1986; 86US-00905146.  
XX  
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.  
XX  
PI Purcell RH, Ticehurst JR, Cohen JL, Emerson SU, Feinstein SM;  
PI Daemer RJ, Gust ID;  
XX  
XX WPI; 1990-075557/10.  
DR N-PSDB; AAQ03512.  
XX  
XX Vaccine against hepatitis A virus infection - comprises novel attenuated  
FT hepatitis A virus strain.  
XX  
PS Claim 1; Fig 1; 18pp; English.  
XX  
XX The attenuated HAV is useful for inducing protective immunity against  
CC HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by  
CC several nucleotide changes distributed throughout the genome, is  
CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is  
CC suitable for use as an HAV vaccine. It is noted that not all the changes  
CC are necessary for attenuation and use as a vaccine. (Updated on 25-MAR-  
CC 2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PI field.)  
CC (Updated on 27-AUG-2003 to correct OS field.) (Updated on 24-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 2227 AA;

Query Match 100.0%; Score 109; DB 2; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20  
|||  
Db 779 VDDPRSEEDKRFESHIECRK 798

RESULT 7  
AAW34074  
ID AAW34074 standard; protein; 2227 AA.  
XX  
AC AAW34074;  
XX  
DT 17-OCT-2003 (revised)  
DT 27-APR-1998 (first entry)  
XX  
DE Hepatitis A virus HM-175 protein sequence.  
XX  
KW HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus; infection;  
KW vaccine.  
XX  
OS Hepatitis A virus; HM-175.

Key	Location/Qualifiers
FT Protein	1..23
FT Protein	/label= VP4
FT Protein	24..245
FT Protein	/label= VP2
FT Protein	246..491
FT Protein	/label= VP3
FT Protein	492..791
FT Protein	/label= VP1
FT Protein	792..980
FT Protein	/label= 2A
FT Protein	981..1087
FT Protein	/label= 2B
FT Protein	1088..1422
FT Protein	/label= 2C
FT Protein	1423..1496
FT Protein	/label= 3A
FT Protein	1497..1519
FT Protein	/label= 3B
FT Protein	1520..1738





CC strain HM-174. The sequence is modified to produce HAV which are adapted  
 CC to growth in the human fibroblast-like cell line MRC-5. The HAV is able  
 CC to propagate in MRC-5 cells and retain appropriate attenuation. It is  
 CC useful as a live vaccine for prophylaxis of hepatitis A in humans and  
 CC other primates  
 XX  
 SQ Sequence 2227 AA;

Query Match 100.0%; Score 109; DB 3; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20  
 |||||  
 Db 779 VDDPRSEEDKRFESHIECRK 798

RESULT 10  
 AAB18608  
 ID AAB18608 standard; protein; 2227 AA.

XX AAB18608;  
 AC  
 DT 15-JAN-2001 (first entry)  
 XX  
 DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.  
 XX  
 KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
 KW P-35 virus.  
 XX

OS Hepatitis A virus.  
 XX  
 PN US6113912-A.  
 XX

PD 05-SEP-2000.  
 XX  
 PF 07-JUN-1995; 95US-00475886.  
 XX

PR 18-SEP-1992; 92US-00947338.  
 PR 17-SEP-1993; 93WO-US008610.  
 PR 17-APR-1995; 95US-00397232.  
 XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
 XX

DR WPI; 2000-586464/55.  
 DR N-PSDB; AAA75477.  
 XX

PT Novel live hepatitis A virus adapted to growth in human fibroblast cell  
 PT line useful as vaccine for protecting humans against hepatitis A virus  
 PT infection, has modified genome compared to wild type.  
 XX

PS Disclosure; Col 67-78; 72pp; English.  
 XX  
 CC The present sequence is derived from passage 35 of a wild type hepatitis  
 CC A virus (HAV) strain HM-174; the resulting virus is designated P-35  
 CC virus. The sequence is modified to produce HAV which are adapted to  
 CC growth in the human fibroblast-like cell line MRC-5. The HAV is able to  
 CC propagate in MRC-5 cells and retain appropriate attenuation. It is useful  
 CC as a live vaccine for prophylaxis of hepatitis A in humans and other  
 CC primates  
 XX

SQ Sequence 2227 AA;  
 Query Match 100.0%; Score 109; DB 3; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20  
 |||||  
 Db 779 VDDPRSEEDKRFESHIECRK 798

RESULT 11  
 ABG31729

ID ABG31729 standard; protein; 2227 AA.

XX  
 AC ABG31729;

XX  
 DT 29-AUG-2003 (revised)

DT 29-NOV-2002 (first entry)

DE Attenuated Hepatitis A virus (HAV) 4380 polypeptide.

XX  
 KW Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast; virucide;  
 KW HAV 4380.

XX  
 OS Hepatitis A virus; strain HM-175.

XX  
 PN US6423318-B1.

XX  
 PD 23-JUL-2002.

XX  
 PF 31-AUG-2000; 2000US-00653499.

XX  
 PR 17-SEP-1993; 93WO-US008610.

PR 17-APR-1995; 95US-00397232.

PR 07-JUN-1995; 95US-00475886.

XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX  
 PI Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;

XX  
 DR WPI; 2002-680946/73.

DR N-PSDB; ABS52789.

PT New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth  
 PT in MRC-5 cells, useful for preparing a vaccine against HAV infection.

XX  
 PS Disclosure; Col 93-104; 71pp; English.

XX  
 CC The invention relates to a polynucleotide which encodes a hepatitis A  
 CC virus (HAV) adapted to growth in MRC-5 cells (human fibroblast-like cell  
 CC line). The polynucleotide is useful for preparing a vaccine against  
 CC hepatitis A virus infection. This sequence represents an attenuated  
 CC hepatitis A virus 4330 polypeptide. (Updated on 29-AUG-2003 to  
 CC standardise OS field)

XX  
 SQ Sequence 2227 AA;

Query Match 100.0%; Score 109; DB 5; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20  
 |||||

Db 779 VDDPRSEEDKRFESHIECRK 798  
 |||||

RESULT 12

ABG31727

ID ABG31727 standard; protein; 2227 AA.

XX  
 AC ABG31727;

XX  
 DT 29-AUG-2003 (revised)

DT 29-NOV-2002 (first entry)

DE Wild-type Hepatitis A virus (HAV) strain HM-175 polypeptide.

XX  
 KW Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast; virucide.

XX  
 OS Hepatitis A virus; strain HM-175.

PN US6423318-B1.  
 XX  
 PD 23-JUL-2002.  
 XX  
 PP 31-AUG-2000; 2000US-00653499.  
 XX  
 PR 17-SEP-1993; 93MO-US008610.  
 PR 17-APR-1995; 95US-00397232.  
 PR 07-JUN-1995; 95US-00475886.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 PI Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
 XX  
 DR WPI; 2002-680946/73.  
 DR N-PSDB; ABS52787.  
 XX  
 PR New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth  
 FT in MRC-5 cells, useful for preparing a vaccine against HAV infection.  
 FT  
 XX  
 PS Disclosure; Fig 6; 71pp; English.  
 XX  
 CC The invention relates to a polynucleotide which encodes a hepatitis A  
 CC virus (HAV) adapted to growth in MRC-5 cells (human fibroblast-like cell  
 CC line). The polynucleotide is useful for preparing a vaccine against  
 CC hepatitis A virus infection. This sequence represents a hepatitis A virus  
 CC strain HM-175 polypeptide. (Updated on 29-AUG-2003 to standardise OS  
 CC field)  
 XX  
 SQ Sequence 2227 AA;  
 Query Match 100.0%; Score 109; DB 5; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VDDPRSEEDKRFESHIECRK 20  
 |||||  
 DB 779 VDDPRSEEDKRFESHIECRK 798

RESULT 13  
 ABG31728  
 ID ABG31728 standard; protein; 2227 AA.  
 AC  
 AC ABG31728;  
 XX  
 XX 29-NOV-2002 (first entry)  
 DE Hepatitis A virus mutant strain HM-175/7 (pHAV/7) polypeptide.  
 XX  
 XX Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast;  
 KW virucide; mutant; pHAV/7; mutain.  
 XX  
 OS Hepatitis A virus; strain HM-175.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 764  
 FT /note= "Wild-type Glu substituted by Val"  
 FT Misc-difference 821  
 FT /note= "Wild-type Asn substituted by Ser"  
 FT Misc-difference 963  
 FT /label= "Wild-type Lys substituted by Arg"  
 FT Misc-difference 1052  
 FT /note= "Wild-type Ala substituted by Val"  
 FT Misc-difference 1062  
 FT /note= "Wild-type Gly substituted by Ala"  
 FT Misc-difference 1118  
 FT /note= "Wild-type Lys substituted by Met"  
 FT Misc-difference 1151  
 FT /note= "Wild-type Glu substituted by Lys"  
 FT Misc-difference 1163

FT /note= "Wild-type Phe substituted by Ser"  
 FT Misc-difference 1277  
 FT /note= "Wild-type Val substituted by Ile"  
 FT Misc-difference 1500  
 FT /note= "Wild-type His substituted by Tyr"  
 FT Misc-difference 1805  
 FT /note= "Wild-type Asp substituted by Asn"  
 FT Misc-difference 1930  
 FT /note= "Wild-type Ser substituted by Thr"  
 XX  
 PN US6423318-B1.  
 XX  
 PD 23-JUL-2002.  
 XX  
 PP 31-AUG-2000; 2000US-00653499.  
 XX  
 PR 17-SEP-1993; 93MO-US008610.  
 PR 17-APR-1995; 95US-00397232.  
 PR 07-JUN-1995; 95US-00475886.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 PI Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
 XX  
 DR WPI; 2002-680946/73.  
 DR N-PSDB; ABS52788.  
 XX  
 PR New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth  
 FT in MRC-5 cells, useful for preparing a vaccine against HAV infection.  
 FT  
 XX  
 PS Example 3; Col 67-78; 71pp; English.  
 XX  
 CC The invention relates to a polynucleotide which encodes a hepatitis A  
 CC virus (HAV) adapted to growth in MRC-5 cells (human fibroblast-like cell  
 CC line). The polynucleotide is useful for preparing a vaccine against  
 CC hepatitis A virus infection. This sequence represents a hepatitis A virus  
 CC mutant strain HM-175/7 (pHAV/7) polypeptide  
 XX  
 SQ Sequence 2227 AA;  
 Query Match 100.0%; Score 109; DB 5; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VDDPRSEEDKRFESHIECRK 20  
 |||||  
 DB 779 VDDPRSEEDKRFESHIECRK 798

RESULT 14  
 ABU08640  
 ID ABU08640 standard; protein; 2227 AA.  
 XX  
 AC ABU08640;  
 XX  
 XX 23-OCT-2003 (revised)  
 DT 03-JUN-2003 (first entry)  
 XX  
 DE Attenuated (pass35) hepatitis A virus strain HM-175.  
 XX  
 KW Hepatitis A virus; HAV; virucide; hepatotropic; antiinflammatory;  
 KW vaccine; MRC-5 cell; hepatitis infection; ds; attenuated virus.  
 XX  
 OS Hepatitis A virus; strain HM-175.  
 XX  
 PN US2002176869-A1.  
 XX  
 PD 28-NOV-2002.  
 XX  
 XX 29-APR-2002; 2002US-00135988.  
 XX  
 PR 18-SEP-1992; 92US-00947338.

PR 17-SEP-1993; 93WO-US008610.  
PR 17-APR-1995; 95US-00397232.  
PR 07-JUN-1995; 95US-00475886.  
PR 31-AUG-2000; 2000US-00653499.  
XX (FUNK/) FUNKHOUSER A W.  
PA (EMER/) EMERSON S U.  
PA (PURC/) PURCELL R H.  
PA (DHON/) D'HONDT E.  
XX  
PI Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
XX  
XX WPI; 2003-352605/02.  
DR N-PSDB; ABX93474.  
XX  
XX New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells,  
PT useful in vaccines for protecting primates against hepatitis infection  
PT and disease.  
XX  
XX Example 3; Fig 6; 70pp; English.  
PS  
XX The invention describes a live hepatitis A virus (HAV) adapted to growth  
CC in MRC-5 cells. The HAV and compositions comprising the HAV are useful as  
CC a vaccine for protecting primates against hepatitis infection and  
CC disease. This is the amino acid sequence of an attenuated human  
CC hepatitis A virus strain HM-175. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 2227 AA;  
Query Match 100.0%; Score 109; DB 6; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 779 VDDPRSEEDKRFESHIECRK 798  
RESULT 15  
ABU08641  
ID ABU08641 standard; protein; 2227 AA.  
AC ABU08641;  
XX  
XX 23-OCT-2003 (revised)  
DT 03-JUN-2003 (first entry)  
XX  
XX Attenuated hepatitis A virus (4380) strain HM-175.  
DE  
XX Hepatitis A virus; HAV; virucide; hepatotropic; antiinflammatory;  
KW vaccine; MRC-5 cell; hepatitis infection; ds; attenuated virus.  
XX  
XX Hepatitis A virus; strain HM-175.  
OS  
XX US2002176869-A1.  
PN  
XX 28-NOV-2002.  
PD  
XX 29-APR-2002; 2002US-00135988.  
PF  
XX 18-SEP-1992; 92US-00947338.  
PR 17-SEP-1993; 93WO-US008610.  
PR 17-APR-1995; 95US-00397232.  
PR 07-JUN-1995; 95US-00475886.  
PR 31-AUG-2000; 2000US-00653499.  
XX  
XX (FUNK/) FUNKHOUSER A W.  
PA (EMER/) EMERSON S U.  
PA (PURC/) PURCELL R H.  
PA (DHON/) D'HONDT E.  
XX  
PI Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
XX

XX WPI; 2003-352605/02.  
DR N-PSDB; ABX93475.  
XX  
PT New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells,  
PT useful in vaccines for protecting primates against hepatitis infection  
PT and disease.  
XX  
PS Disclosure; Page 45-51; 70pp; English.  
XX  
XX The invention describes a live hepatitis A virus (HAV) adapted to growth  
CC in MRC-5 cells. The HAV and compositions comprising the HAV are useful as  
CC a vaccine for protecting primates against hepatitis infection and  
CC disease. This is the amino acid sequence of an attenuated human  
CC hepatitis A virus (4380) strain HM-175. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 2227 AA;  
Query Match 100.0%; Score 109; DB 6; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 779 VDDPRSEEDKRFESHIECRK 798  
RESULT 16  
ABU08639  
ID ABU08639 standard; protein; 2227 AA.  
XX  
AC ABU08639;  
XX  
DT 23-OCT-2003 (revised)  
DT 03-JUN-2003 (first entry)  
XX  
XX Wild type human hepatitis A virus strain HM-175.  
DE  
XX Hepatitis A virus; HAV; virucide; hepatotropic; antiinflammatory;  
KW vaccine; MRC-5 cell; hepatitis infection.  
XX  
XX Hepatitis A virus; strain HM-175.  
OS  
XX US2002176869-A1.  
PN  
XX 28-NOV-2002.  
PD  
XX 29-APR-2002; 2002US-00135988.  
PF  
XX 18-SEP-1992; 92US-00947338.  
PR 17-SEP-1993; 93WO-US008610.  
PR 17-APR-1995; 95US-00397232.  
PR 07-JUN-1995; 95US-00475886.  
PR 31-AUG-2000; 2000US-00653499.  
XX  
XX (FUNK/) FUNKHOUSER A W.  
PA (EMER/) EMERSON S U.  
PA (PURC/) PURCELL R H.  
PA (DHON/) D'HONDT E.  
XX  
XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
PI  
XX WPI; 2003-352605/02.  
DR N-PSDB; ABX93473.  
XX  
XX New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells,  
PT useful in vaccines for protecting primates against hepatitis infection  
PT and disease.  
XX  
PS Disclosure; Fig 6; 70pp; English.  
XX  
XX The invention describes a live hepatitis A virus (HAV) adapted to growth  
CC

CC in MRC-5 cells. The HAV and compositions comprising the HAV are useful as  
CC a vaccine for protecting primates against hepatitis infection and  
CC disease. This is the amino acid sequence of wild type human hepatitis A  
CC virus strain HM-175. (Updated on 23-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 2227 AA;

Query Match 100.0%; Score 109; DB 6; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

OY 1 VDDPRSEEDKRFESHIECK 20  
|||||:|||||  
Db 779 VDDPRSEEDKRFESHIECK 798

#### RESULT 17

AAP50230  
ID AAP50230 standard; protein; 366 AA.

XX AC AAP50230;

XX DT 28-NOV-1991 (first entry)

XX DE Sequence of hepatitis A virus (HAV) surface protein (VP-1).

XX KW Hepatitis A virus vaccine; immunisation; monoclonal antibody;

XX KM diagnostic assay.

XX OS Hepatitis A virus.

XX PN EP138704-A.

XX PD 24-APR-1985.

XX PF 09-OCT-1984; 84EP-00402025.

XX PR 14-OCT-1983; 83US-00541836.

XX PR 02-MAR-1984; 84US-00585942.

XX PA (MERI ) MERCK & CO INC.

XX PI Hughes JV, Scolnick EM, Tomassini JE;

XX DR WPI: 1985-100818/17.

XX DR N-PSDB; AAN50274.

XX PT New hepatitis A virus surface protein - useful for binding to  
PT neutralising antibodies to the virus.

XX PS Claim 21; Page 46-48; 49pp; English.

XX CC VP1 is isolated by solubilisation of the intact virus in an aq. antonic  
CC surfactant and a reducing agent. The viral proteins are sepd. and the  
CC protein of molecular wt. 33000 daltons is sepd

XX SQ Sequence 366 AA;

Query Match 97.2%; Score 106; DB 1; Length 366;  
Best Local Similarity 95.0%; Pred. No. 6.8e-08;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VDDPRSEEDKRFESHIECK 20  
|||||:|||||  
Db 288 VDDPRSEEDKRFESHIECK 307

#### RESULT 18

AAP50116  
ID AAP50116 standard; protein; 993 AA.

XX AC AAP50116;

XX XX

DT 24-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 30-SEP-1991 (first entry)  
XX Sequence of Hepatitis A virus (HAV) immunogenic peptides VP-1, VP-2, VP-3  
DE and VP-4.

XX KW Antigenic protein; immunogen; vaccine.

XX OS Hepatitis A virus; (strain CR326).

XX PN EP154587-A.

XX PD 11-SEP-1985.

XX PF 27-FEB-1985; 85EP-00400369.

XX PR 02-MAR-1984; 84US-00585818.

XX PA (MERI ) MERCK & CO INC.

XX PI Linemeyer DL, Menke JG, Rueben RG, Mitra SW;

XX DR WPI: 1985-224964/37.

XX DR N-PSDB; AAN50139.

XX PT New nucleotide sequences coding for hepatitis A virus antigens - useful  
PT for eliciting normal immune response and in vaccines for protecting  
PT against the virus.

XX PS Example; Page 11-17; 32pp; English.

XX CC Within the sequence in AAN50139 is encoded the information necessary to  
CC make the antigenic proteins of HAV. The sequences encoding for the  
CC structural proteins begin at base 403. The key sub-unit sequences within  
CC VP-1, designated Sequences I, II, III, IV, and V, start, respectively at  
CC 1882, 1983, 1999, 2146, 2347. Other nucleotide sequences which are  
CC valuable as encoding antigenic proteins are the sequences from base 1749  
CC to base 2722; from base 1487 to base 2980 and from base 1644 to base  
CC 2722. The sequence from base 1749 to base 2722 is esp. valuable as a  
CC vector for producing antigen protein. Sequences II-V are claimed. X in  
CC AAP50116 denotes the translation of a stop codon. (Updated on 25-MAR-2003  
CC to correct PI field.) (Updated on 24-OCT-2003 to standardise OS field)

XX SQ Sequence 993 AA;

Query Match 97.2%; Score 106; DB 1; Length 993;  
Best Local Similarity 95.0%; Pred. No. 2e-07; Indels 0; Gaps 0;  
Matches 19; Conservative 1; Mismatches 0;

OY 1 VDDPRSEEDKRFESHIECK 20  
|||||:|||||  
Db 915 VDDPRSEEDKRFESHIECK 934

#### RESULT 19

AAP50231  
ID AAP50231 standard; protein; 993 AA.

XX AC AAP50231;

XX DT 28-NOV-1991 (first entry)

XX DE Sequence encoded by partial sequence of hepatitis A virus (HAV),  
DE including surface protein (VP-1).

XX KW Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
KW diagnostic assay.

XX OS Hepatitis A virus.

XX FH Key Location/Qualifiers

FT Protein 628..993

FT /note= "claimed; X denotes translated stop codons and  
FT unspecified triplets"  
XX  
XX EPI38704-A.  
XX  
XX 24-APR-1985.  
XX PD  
XX 09-OCT-1984; 84EP-00402025.  
XX PF  
XX 14-OCT-1983; 83US-00541836.  
XX PR  
XX 02-MAR-1984; 84US-00585942.  
XX PR  
XX (MERI ) MERCK & CO INC.  
XX PA  
XX  
XX Hughes JV, Scolnick EM, Tomassini JE;  
XX WPI; 1985-100818/17.  
XX DR N-PSDB; AAN50274.  
XX  
XX New hepatitis A virus surface protein - useful for binding to  
XX neutralising antibodies to the virus.  
XX PT  
XX  
XX Disclosure; Page 17-23; 49pp; English.  
XX PS  
XX VP1 is isolated by solubilisation of the intact virus in an aq. anionic  
XX surfactant and a reducing agent. The viral proteins are sepd. and the  
XX protein of molecular wt. 33000 Daltons is sepd  
XX CC  
XX Sequence 993 AA;  
XX  
XX Query Match 97.2%; Score 106; DB 1; Length 993;  
XX Best Local Similarity 95.0%; Pred. No. 2e-07; Mismatches 0; Indels 0; Gaps 0;  
XX Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 VDDPRSEEDKRFESHIECRK 20  
XX |||||:|||||  
XX 915 VDDPRSEEDRRFESHIECRK 934  
XX  
XX  
XX RESULT 20  
XX AAE19899  
XX ID AAE19899 standard; protein; 2227 AA.  
XX AC AAE19899;  
XX  
XX 18-JUN-2002 (first entry)  
XX DT  
XX  
XX Hepatitis A virus (HAV) protein.  
XX DE  
XX Hepatitis A virus; HAV; infection; virucide; fungicide; antibacterial;  
XX KW cytostatic; immunostimulant; vaccine; ribavirin; immune response; cancer.  
XX  
XX Hepatitis A virus.  
XX OS  
XX WO200213855-A2.  
XX PN  
XX 21-FEB-2002.  
XX PD  
XX  
XX 15-AUG-2001; 2001WO-IB001808.  
XX PF  
XX  
XX 17-AUG-2000; 2000US-0225767P.  
XX PR  
XX 29-AUG-2000; 2000US-0229175P.  
XX PR  
XX 03-NOV-2000; 2000US-00705547.  
XX  
XX (TRIP-) TRIPEP AB.  
XX PA  
XX Sallberg M, Hultgren C;  
XX PI  
XX WPI; 2002-241837/29.  
XX DR N-PSDB; AAD31766.  
XX  
XX Vaccine compositions for treating and preventing disease, preferably  
XX hepatitis C virus infection, comprises ribavirin and antigen that has

FT epitope present in hepatitis C virus.  
XX  
XX Claim 11; Page 82-87; 120pp; English.  
XX PS  
XX The invention relates to a composition comprising ribavirin and an  
XX antigen preferably non structural 3 protein (NS3)/4A fragment of  
XX hepatitis C virus (HCV) genome or a peptide or nucleic acid of HCV  
XX sequence. The composition is useful for enhancing an immune response to a  
XX hepatitis C antigen in humans, domestic, sport or pet species and as  
XX vaccines for treating and preventing HCV infections. The composition is  
XX also useful for treating viral, bacterial, fungal diseases and cancer.  
XX CC The present sequence is hepatitis A virus (HAV) protein  
XX  
XX Sequence 2227 AA;  
XX  
XX Query Match 97.2%; Score 106; DB 5; Length 2227;  
XX Best Local Similarity 95.0%; Pred. No. 4.8e-07;  
XX Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 VDDPRSEEDKRFESHIECRK 20  
XX |||||:|||||  
XX 779 VDDPRSEEDRRFESHIECRK 798  
XX  
XX  
XX RESULT 21  
XX ABW00350  
XX ID ABW00350 standard; protein; 2227 AA.  
XX AC ABW00350;  
XX  
XX 15-JAN-2004 (first entry)  
XX DT  
XX Hepatitis A virus protein.  
XX DE  
XX Ribavirin; vaccine; immune response; infection; therapy; immunostimulant;  
XX KW virucide.  
XX  
XX Hepatitis A virus.  
XX OS  
XX US2002136740-A1.  
XX PN  
XX 26-SEP-2002.  
XX PD  
XX  
XX 15-AUG-2001; 2001US-00929955.  
XX PF  
XX 17-AUG-2000; 2000US-0225767P.  
XX PR  
XX 29-AUG-2000; 2000US-0229175P.  
XX PR  
XX (SALL/) SALLBERG M.  
XX PA (HULT/) HULTGREN C.  
XX  
XX Sallberg M, Hultgren C;  
XX PI  
XX WPI; 2003-764978/72.  
XX DR N-PSDB; AAD60867.  
XX  
XX Vaccine compositions for treating and preventing disease, preferably  
XX hepatitis C virus infection, comprises ribavirin and antigen that has  
XX epitope present in hepatitis C virus.  
XX  
XX Claim 11; Page 45-51; Opp; English.  
XX PS  
XX The invention relates to a composition comprising ribavirin and an  
XX antigen, where the antigen is derived from a hepatitis virus. The vaccine  
XX is useful in enhancing the immune response to a hepatitis C antigen where  
XX the composition is delivered to an animal identified as requiring an  
XX enhanced immune response. The vaccine is useful in the treatment and  
XX prevention of hepatitis C infection. The present sequence is Hepatitis A  
XX virus protein  
XX  
XX Sequence 2227 AA;  
XX  
XX Query Match 97.2%; Score 106; DB 7; Length 2227;  
XX

Best Local Similarity 95.0%; Pred. No. 4.8e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRPESHIECRK 20  
|||||:|||||  
Db 779 VDDPRSEEDKRPESHIECRK 798

RESULT 22  
ID AAP60066 standard; protein; 2227 AA.  
XX AAP60066;  
XX  
XX 25-MAR-2003 (revised)  
DT 26-JUN-1991 (first entry)  
XX  
DE Sequence of viral L434 polypeptide encoded by the complete nucleotide  
DE sequence of the HAV genome.  
XX  
XX Diagnosis; vaccine; passive immunotherapy.  
XX  
OS Hepatitis A virus.  
XX  
FH Key Location/Qualifiers  
FT Region 1..245  
FT /label= P1.1A  
FT Region 246..491  
FT /label= 1B  
FT Region 492..836  
FT /label= 1C  
FT Region 837..980  
FT /label= P2.2A  
FT Region 991..1076  
FT /label= 2B  
FT Region 1077..1422  
FT /label= 2C  
FT Region 1423..1484  
FT /label= P3.3A  
FT Region 1485..1507  
FT /label= 3B  
FT Region 1508..1678  
FT /label= 3C  
FT Region 1679..2227  
FT /label= 3D  
XX  
XX EP199480-A.  
XX  
XX 29-OCT-1986.  
XX  
XX 03-APR-1986; 86EP-00302465.  
XX  
XX 03-APR-1985; 85US-00719329.  
XX  
XX (CHIR ) CHIRON CORP.  
XX  
XX Dina D, Potter SJ, Vannest GA, Caput D;  
PI WPI; 1986-286213/44.  
XX  
XX N-PSDB; AAP60080.  
XX  
XX Hepatitis A virus nucleotide sequence and polypeptide - and use in prodn.  
FT of vaccines and diagnostic probes.  
XX  
XX Claim 5; Fig 1; 18pp; English.  
XX  
XX AAP60080 and oligonucleotide fragments are useful in detection of  
CC hepatitis A virus; transformed hosts may be used for expression of  
CC polypeptides and fragments useful in vaccines without risk of infection  
CC by the virus or in prodn. of particles which are capable of inducing  
CC immunocompetent B cells for passive immunotherapy. Pref. epitope is  
CC derived from AAs 445-657 or 792-848 of the HAV polypeptide sequence  
CC (AAP60066). (Updated on 25-MAR-2003 to correct PA field.)

XX SQ Sequence 2227 AA;  
Query Match 79.8%; Score 87; DB 1; Length 2227;  
Best Local Similarity 80.0%; Pred. No. 0.0004;  
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRPESHIECRK 20  
|||||:|||||  
Db 779 VDDPRSEEDKRPESHIECRK 798

RESULT 23  
ID AAR15629 standard; protein; 839 AA.  
XX AAR15629;  
XX  
XX 24-OCT-2003 (revised)  
DT 27-AUG-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 17-DEC-2001 (revised)  
DT 17-MAR-1992 (first entry)  
XX  
DE Capsid region of cyno-HAV isolate CY-145.  
XX  
XX Hepatitis A virus; cynomolgus; HAV; monkey; vaccine; macaque.  
XX  
OS Viruses.  
XX  
FH Key Location/Qualifiers  
FT Protein 1..245  
FT /label= VP0  
FT Cleavage-site 245..246  
FT Protein 246..491  
FT /label= VP3  
FT Active-site 315  
FT Cleavage-site 491..492  
FT Protein 492..791  
FT /label= VP1  
FT Active-site 593  
FT Cleavage-site 791..792  
FT Protein 792  
FT /label= P2  
FT /note= "incomplete"  
XX  
XX USN7678828-N.  
XX  
XX 12-NOV-1991.  
XX  
XX 03-APR-1991; 91US-00678828.  
XX  
XX 03-APR-1991; 91US-00678828.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
XX  
XX Hainan OV, Margolis HS, Robertson BH, Brinton MH, Ebert JW;  
PI WPI; 1991-376737/51.  
XX  
XX N-PSDB; AAQ15180.  
XX  
XX Hepatitis A virus isolates and DNA - used to prepare vaccines for  
FT preventing hepatitis A virus infection.  
XX  
XX Disclosure; Fig 3; 23pp; English.  
XX  
XX The sequence was deduced from the nucleotide sequence obtd. by PCR  
CC amplification of cyno-HAV viral RNA obtd. from the stool of a cynomolgus  
CC monkey with serologically and histologically confirmed spontaneous  
CC hepatitis A. The sequence differs from the human HAV isolate HMI75  
CC (Cohen, J.I., et al. (1987) Proc. Natl. Acad. Sci. USA 84, 2497-2501),  
CC mainly in the VP3 and VP1 proteins. The Gln-Val pair at the VP3-VP1  
CC cleavage site in the human isolate is replaced by a Gln-Thr pair in the

CC cyno-HAV. The other two cleavage sites are the same. Two residues have  
CC been identified as part of the immuno- dominant region (see feature  
CC table) and are different to those in the same position in human HAV. The  
CC protein and peptides derived from it can be used in the prepn. of  
CC vaccines for the prevention of HAV infection. See also AAR15056. (Note:  
CC Revised entry submitted to correct the patent number format of US  
CC Government-owned NTRIS applications to prevent clashes with ongoing US  
CC granted patent numbers. For further information please visit the Derwent  
CC web site at [www.derwent.com/dwpi/updates/ntis.us.html](http://www.derwent.com/dwpi/updates/ntis.us.html).) (Updated on 25-  
CC MAR-2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS  
CC field.) (Updated on 24-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 839 AA;

Query Match 74.3%; Score 81; DB 2; Length 839;  
Best Local Similarity 82.4%; Pred. No. 0.0012;  
Matches 14; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIE 17  
DB 779 VDDPTDEDRFESHIE 795  
|||||:|||||

RESULT 24  
AAW42920  
ID AAW42920 standard; peptide; 20 AA.  
XX  
AC AAW42920;

DT 28-APR-1998 (first entry)

XX Immunogenic Hepatitis A virus peptide YK-1313.

DE Immunogenic peptide; immunogenic epitope; VP1 protein; immune response;  
KW antibody.

XX Synthetic.

OS Hepatitis A virus.

XX WO9740147-A1.

PD 30-OCT-1997.

PF 18-APR-1997; 97WO-US006891.

PR 19-APR-1996; 96US-0015644P.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

PI WPI; 1997-535831/49.

XX Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
PT response to HAV in a mammal or to detect the presence of antibodies  
PT against HAV in a mammal.

PS Claim 13; Page 111; 140pp; English.

XX Peptides AAW42906-21 are immunogenic peptides corresponding to  
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
CC substantially similar to a portion of the amino acid sequence of the VP1  
CC protein of HAV corresponding to amino acids 492-791. The present peptide  
CC is derived from amino acids 772-791, and has a reactivity of 37.5% with  
CC acute sera. Compositions containing the peptides can be used to induce an  
CC immune response to HAV in a mammal. The peptides can also be used to  
CC detect the presence of antibodies against HAV in mammalian serum. The  
CC peptides can also be used to make an antibody against HAV by  
CC administering the peptide to a mammal

XX Sequence 20 AA;

Query Match 63.3%; Score 69; DB 2; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.0014;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRF 13  
DB 8 VDDPASEEDKRF 20  
|||||:|||||

RESULT 25

AAW69437  
ID AAW69437 standard; peptide; 21 AA.

XX AAW69437;

DT 20-APR-2001 (first entry)

XX Synthetic HAV VP1 peptide, SEQ ID NO: 37.

XX Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
KW antigen; major structural capsid polypeptide; HAV antibody detection.

XX Hepatitis A virus.

OS Synthetic.

XX WO200105824-A2.

PD 25-JAN-2001.

PF 14-JUL-2000; 2000WO-US019267.

PR 15-JUL-1999; 99US-0144412P.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

PI WPI; 2001-112681/12.

XX Synthetic peptides used as antigen sources for enzyme immunoassays  
PT detecting anti-hepatitis A virus and as vaccines.

PS Claim 10; Page 91; 130pp; English.

XX The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of antibodies against HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IGM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy end  
CC of the peptides enhances the IGM antibody reactivity

XX Sequence 21 AA;

Query Match 63.3%; Score 69; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.0015;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRF 13  
DB 8 VDDPSEEDKRF 20  
|||||:|||||

RESULT 26  
AAP90612



ID AAP90612 standard; protein; 300 AA.  
 AC AAP90612;  
 DT 25-MAR-2003 (revised)  
 DT 20-JUL-2000 (revised)  
 DT 31-OCT-1989 (first entry)  
 XX  
 DE Portion of structural protein of VP1 of hepatitis A virus strain HM175.  
 XX  
 KW Hepatitis A virus; VP1; HM175 strain; DNA; immunogen; conjugate;  
 KW vaccines; VP3.  
 OS Hepatitis A virus.  
 XX  
 XX WO8906658-A.  
 PN  
 XX 27-JUL-1989.  
 PD  
 PF 13-JAN-1989; 89WO-US000097.  
 XX  
 XX 13-JAN-1988; 88US-00143395.  
 PR  
 XX (UYN-C) UNIV OF N CAROLINA.  
 PA  
 XX Lemon SM;  
 PI  
 XX WPI; 1989-233826/32.  
 DR  
 DR N-PSDB; AAN90399.  
 XX  
 XX Synthetic immunogenic peptide(s) corresp. to hepatitis A virus - derived  
 PT from VP3 structural protein and used for vaccine prodn. and diagnosis.  
 XX  
 PS Disclosure; Fig 1; 29pp; English.  
 XX  
 XX Portion of VP1 (see AAN90399; depicting bases 2197-3106) structural  
 CC protein of HM175 strain hepatitis A virus. The patent claims an  
 CC immunogenic peptide corresp. to a region of VP3, and this can form an  
 CC immunogenic conjugate with eg a portion of VP1 (contg. Ser-102). For VP3  
 CC see AAN90400, AAP90613. (Revised entry issued to correct the sequence  
 CC analysis breakdown.) (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 300 AA;  
 Query Match 63.3%; Score 69; DB 1; Length 300;  
 Best Local Similarity 100.0%; Pred. No. 0.027;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VDDPRSEEDKQFE 13  
 |||||  
 Db 288 VDDPRSEEDKQFE 300  
 RESULT 27  
 AEM69971  
 ID AEM69971 standard; protein; 409 AA.  
 XX  
 AC AEM69971;  
 XX  
 XX 20-NOV-2003 (first entry)  
 DT  
 XX  
 DE Photorhabdus luminescens protein sequence #3068.  
 XX  
 XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 XX WO200294867-A2.  
 PN  
 XX 28-NOV-2002.  
 PD

XX 07-FEB-2002; 2002WO-IB003040.  
 PF  
 XX 07-FEB-2001; 2001FR-00001659.  
 PR  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 XX  
 DR WPI; 2003-148459/14.  
 XX  
 XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 PT  
 XX Claim 2; SEQ ID NO 3068; 1205pp; French.  
 PS  
 XX The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins  
 XX  
 SQ Sequence 409 AA;  
 Query Match 45.9%; Score 50; DB 6; Length 409;  
 Best Local Similarity 57.1%; Pred. No. 31;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 7 EEDKRFESHIECRK 20  
 |. ||| :. |||  
 Db 355 EDQKRTATVECEK 368  
 RESULT 28  
 AAG57588  
 ID AAG57588 standard; protein; 183 AA.  
 XX  
 AC AAG57588;  
 XX  
 XX 18-OCT-2000 (first entry)  
 DT  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 74230.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 XX EP1033405-A2.  
 PN  
 XX 06-SEP-2000.  
 PD  
 XX 25-FEB-2000; 2000EP-00301439.  
 PF  
 XX

PR	25-FEB-1999;	99US-0121825P.	PR	13-JUL-1999;	99US-0143542P.
PR	05-MAR-1999;	99US-0123180P.	PR	14-JUL-1999;	99US-0143624P.
PR	09-MAR-1999;	99US-0123548P.	PR	15-JUL-1999;	99US-0144005P.
PR	23-MAR-1999;	99US-0125788P.	PR	16-JUL-1999;	99US-0144085P.
PR	28-MAR-1999;	99US-0126264P.	PR	16-JUL-1999;	99US-0144085P.
PR	29-MAR-1999;	99US-0126785P.	PR	19-JUL-1999;	99US-0144325P.
PR	01-APR-1999;	99US-0127462P.	PR	19-JUL-1999;	99US-0144331P.
PR	06-APR-1999;	99US-0128234P.	PR	19-JUL-1999;	99US-0144332P.
PR	08-APR-1999;	99US-0128714P.	PR	19-JUL-1999;	99US-0144333P.
PR	16-APR-1999;	99US-0129845P.	PR	19-JUL-1999;	99US-0144334P.
PR	19-APR-1999;	99US-0130077P.	PR	19-JUL-1999;	99US-0144335P.
PR	21-APR-1999;	99US-0130449P.	PR	20-JUL-1999;	99US-0144352P.
PR	23-APR-1999;	99US-0130510P.	PR	20-JUL-1999;	99US-0144632P.
PR	28-APR-1999;	99US-0130891P.	PR	20-JUL-1999;	99US-0144884P.
PR	28-APR-1999;	99US-0131449P.	PR	21-JUL-1999;	99US-0144814P.
PR	30-APR-1999;	99US-0132048P.	PR	21-JUL-1999;	99US-0145086P.
PR	30-APR-1999;	99US-0132407P.	PR	21-JUL-1999;	99US-0145088P.
PR	04-MAY-1999;	99US-0132484P.	PR	22-JUL-1999;	99US-0145085P.
PR	05-MAY-1999;	99US-0132485P.	PR	22-JUL-1999;	99US-0145087P.
PR	06-MAY-1999;	99US-0132486P.	PR	22-JUL-1999;	99US-0145089P.
PR	07-MAY-1999;	99US-0132487P.	PR	22-JUL-1999;	99US-0145089P.
PR	07-MAY-1999;	99US-0132863P.	PR	23-JUL-1999;	99US-0145192P.
PR	11-MAY-1999;	99US-0134256P.	PR	23-JUL-1999;	99US-0145145P.
PR	14-MAY-1999;	99US-0134218P.	PR	23-JUL-1999;	99US-0145218P.
PR	14-MAY-1999;	99US-0134219P.	PR	23-JUL-1999;	99US-0145224P.
PR	14-MAY-1999;	99US-0134219P.	PR	26-JUL-1999;	99US-0145276P.
PR	14-MAY-1999;	99US-0134219P.	PR	27-JUL-1999;	99US-0145913P.
PR	14-MAY-1999;	99US-0134370P.	PR	27-JUL-1999;	99US-0145913P.
PR	18-MAY-1999;	99US-0134768P.	PR	27-JUL-1999;	99US-0145918P.
PR	19-MAY-1999;	99US-0134941P.	PR	27-JUL-1999;	99US-0145919P.
PR	20-MAY-1999;	99US-0135124P.	PR	28-JUL-1999;	99US-0145951P.
PR	21-MAY-1999;	99US-0135353P.	PR	28-JUL-1999;	99US-0145951P.
PR	24-MAY-1999;	99US-0135629P.	PR	02-AUG-1999;	99US-0146386P.
PR	25-MAY-1999;	99US-0136021P.	PR	02-AUG-1999;	99US-0146388P.
PR	27-MAY-1999;	99US-0136392P.	PR	02-AUG-1999;	99US-0146389P.
PR	28-MAY-1999;	99US-0136782P.	PR	03-AUG-1999;	99US-0147038P.
PR	01-JUN-1999;	99US-0137222P.	PR	04-AUG-1999;	99US-0147204P.
PR	01-JUN-1999;	99US-0137528P.	PR	04-AUG-1999;	99US-0147302P.
PR	04-JUN-1999;	99US-0137724P.	PR	05-AUG-1999;	99US-0147192P.
PR	07-JUN-1999;	99US-0138094P.	PR	05-AUG-1999;	99US-0147260P.
PR	08-JUN-1999;	99US-0138094P.	PR	06-AUG-1999;	99US-0147303P.
PR	10-JUN-1999;	99US-0138540P.	PR	06-AUG-1999;	99US-0147416P.
PR	10-JUN-1999;	99US-0138847P.	PR	06-AUG-1999;	99US-0147416P.
PR	14-JUN-1999;	99US-0139119P.	PR	09-AUG-1999;	99US-0147493P.
PR	16-JUN-1999;	99US-0139452P.	PR	09-AUG-1999;	99US-0147935P.
PR	16-JUN-1999;	99US-0139453P.	PR	09-AUG-1999;	99US-0147935P.
PR	17-JUN-1999;	99US-0139452P.	PR	10-AUG-1999;	99US-0148171P.
PR	17-JUN-1999;	99US-0139452P.	PR	11-AUG-1999;	99US-0148319P.
PR	18-JUN-1999;	99US-0139454P.	PR	12-AUG-1999;	99US-0148341P.
PR	18-JUN-1999;	99US-0139455P.	PR	13-AUG-1999;	99US-0148656P.
PR	18-JUN-1999;	99US-0139456P.	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	99US-0139457P.	PR	16-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	99US			

XX	AA	AGS7587;	
XX	AC		
XX	XX		
XX	DT	18-OCT-2000 (first entry)	
XX	DE		
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 74229.	
XX	XX		
XX	KW	Protein identification; signal transduction pathway; metabolic pathway;	
XX	KW	hybridisation assay; Genetic mapping; gene expression control; promoter;	
XX	KW	termination sequence.	
XX	XX		
OS	OS	Arabidopsis thaliana.	
XX	XX		
PN	PN	EP1033405-A2.	
XX	XX		
XX	PD	06-SEP-2000.	
XX	XX		
PF	PF	25-FEB-2000; 2000EP-00301439.	
XX	XX		
PR	PR	25-FEB-1999; 99US-0121825P.	
PR	PR	05-MAR-1999; 99US-0123180P.	
PR	PR	09-MAR-1999; 99US-0123548P.	
PR	PR	23-MAR-1999; 99US-0125788P.	
PR	PR	23-MAR-1999; 99US-0126264P.	
PR	PR	29-MAR-1999; 99US-0126785P.	
PR	PR	01-APR-1999; 99US-0127462P.	

19-JUL-1999;	PR	9905-0144332P;
19-JUL-1999;	PR	9905-0144333P;
19-JUL-1999;	PR	9905-0144334P;
19-JUL-1999;	PR	9905-0144335P;
20-JUL-1999;	PR	9905-0144332P;
20-JUL-1999;	PR	9905-0144335P;
20-JUL-1999;	PR	9905-0144894P;
21-JUL-1999;	PR	9905-0144814P;
21-JUL-1999;	PR	9905-0145066P;
21-JUL-1999;	PR	9905-0145085P;
22-JUL-1999;	PR	9905-0145085P;
22-JUL-1999;	PR	9905-0145087P;
22-JUL-1999;	PR	9905-0145089P;
22-JUL-1999;	PR	9905-0145192P;
23-JUL-1999;	PR	9905-0145145P;
23-JUL-1999;	PR	9905-0145218P;
23-JUL-1999;	PR	9905-0145224P;
26-JUL-1999;	PR	9905-0145276P;
27-JUL-1999;	PR	9905-0145913P;
27-JUL-1999;	PR	9905-0145918P;
28-JUL-1999;	PR	9905-0145951P;
02-AUG-1999;	PR	9905-0146386P;
02-AUG-1999;	PR	9905-0146388P;
03-AUG-1999;	PR	9905-0147038P;
04-AUG-1999;	PR	9905-0147204P;
04-AUG-1999;	PR	9905-0147302P;
05-AUG-1999;	PR	9905-0147182P;
06-AUG-1999;	PR	9905-0147260P;
06-AUG-1999;	PR	9905-0147303P;
09-AUG-1999;	PR	9905-0147416P;
09-AUG-1999;	PR	9905-0147493P;
10-AUG-1999;	PR	9905-0147935P;
11-AUG-1999;	PR	9905-0148171P;
12-AUG-1999;	PR	9905-0148319P;
13-AUG-1999;	PR	9905-0148341P;
13-AUG-1999;	PR	9905-0148565P;
16-AUG-1999;	PR	9905-0148684P;
17-AUG-1999;	PR	9905-0149368P;
18-AUG-1999;	PR	9905-0149426P;
18-AUG-1999;	PR	9905-0149722P;
20-AUG-1999;	PR	9905-0149723P;
20-AUG-1999;	PR	9905-0149923P;
23-AUG-1999;	PR	9905-0149902P;
23-AUG-1999;	PR	9905-0149930P;
25-AUG-1999;	PR	9905-0150566P;
26-AUG-1999;	PR	9905-0150884P;
27-AUG-1999;	PR	9905-0151065P;
27-AUG-1999;	PR	9905-0151066P;
27-AUG-1999;	PR	9905-0151080P;
30-AUG-1999;	PR	9905-0151303P;
31-AUG-1999;	PR	9905-0151438P;
01-SEP-1999;	PR	9905-0151930P;
01-SEP-1999;	PR	9905-0152363P;
10-SEP-1999;	PR	9905-0153070P;
13-SEP-1999;	PR	9905-0153759P;
15-SEP-1999;	PR	9905-0154018P;
16-SEP-1999;	PR	9905-0154039P;
22-SEP-1999;	PR	9905-0155139P;
23-SEP-1999;	PR	9905-0155486P;
24-SEP-1999;	PR	9905-0155659P;
28-SEP-1999;	PR	9905-0156458P;
29-SEP-1999;	PR	9905-0156596P;
04-OCT-1999;	PR	9905-0157117P;
05-OCT-1999;	PR	9905-0157753P;
06-OCT-1999;	PR	9905-0157865P;
08-OCT-1999;	PR	9905-0158025P;
08-OCT-1999;	PR	9905-0158232P;
12-OCT-1999;	PR	9905-0158369P;
13-OCT-1999;	PR	9905-0159293P;

PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 30-JUN-1999; 99US-0140991P.  
PR 01-JUL-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 20-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144684P.

PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 02-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147202P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149829P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158312P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.

```
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 42.7%; Score 46.5; DB 3; Length 200;
Best Local Similarity 50.0%; Pred. No. 49;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 1 VDDPRSEED-KREESHIECR 19
Db 157 IDDSLQETKRFVSHVEAR 176

Search completed: May 11, 2004, 13:37:21
Job time : 52 secs
```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:30:01 ; Search time 10.5 Seconds  
(without alignments)  
183.222 Million cell updates/sec

Title: US-09-171-432a-38

Perfect score: 109

Sequence: 1 VDDPRSEDKRPFESHCCK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR 78.\*

1: p1r1.\*

2: p1r2.\*

3: p1r3.\*

4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	56	PQ0431	genome polypotein
2	109	100.0	319	JH0135	genome polypotein
3	109	100.0	1358	A03905	genome polypotein
4	109	100.0	2227	GNNYHM	genome polypotein
5	109	100.0	2227	GNNYMK	genome polypotein
6	109	100.0	2227	GNNYHB	genome polypotein
7	106	97.2	56	PQ0427	genome polypotein
8	106	97.2	56	PQ0429	genome polypotein
9	106	97.2	341	S04137	genome polypotein
10	106	97.2	852	GNNYHA	genome polypotein
11	106	97.2	2227	GNNYHR	genome polypotein
12	99	90.8	56	PQ0430	genome polypotein
13	96	88.1	56	PQ0432	genome polypotein
14	95	87.2	56	PQ0428	genome polypotein
15	91	83.5	56	PQ0434	genome polypotein
16	84	77.1	56	PQ0437	genome polypotein
17	84	77.1	56	PQ0436	genome polypotein
18	84	77.1	2230	GNNYSA	genome polypotein
19	81	74.3	55	PQ0433	genome polypotein
20	81	74.3	55	PQ0435	genome polypotein
21	81	74.3	839	GNNY52	genome polypotein
22	60	55.0	300	A33327	genome polypotein
23	51	46.8	1548	T25808	hypothetical prote
24	50.5	46.3	754	T25551	hypothetical prote
25	50	45.9	1663	1 C3RT	complement C3 prec
26	50	45.9	1666	1 C3GP	complement C3 - Af
27	48	44.0	322	A32329	legumin B - fava b
28	48	44.0	564	2 S37241	hypothetical prote
29	46.5	42.7	454	2 B84599	hypothetical prote

30	46	42.2	242	2	E83077	conserved hypothet
31	46	42.2	308	2	T17524	hypothetical prote
32	46	42.2	1663	1	C3MS	complement C3 prec
33	45.5	41.7	650	2	T36419	hypothetical prote
34	45	41.3	209	2	T20155	hypothetical prote
35	45	41.3	480	2	T20154	hypothetical prote
36	45	41.3	586	1	E69314	replication licens
37	45	41.3	741	2	S73827	hypothetical prote
38	44	40.4	222	2	E71507	probable L4 riboso
39	44	40.4	234	2	B83065	conserved hypothet
40	44	40.4	433	2	A23850	phosphoryruvate hy
41	44	40.4	490	2	C84091	hypothetical prote
42	44	40.4	499	2	T04730	cytochrome P450 ho
43	44	40.4	593	2	A81021	para-aminobenzoate
44	44	40.4	598	2	D81965	probable para-amin
45	44	40.4	782	2	G96598	hypothetical prote
46	44	40.4	957	2	T50789	hypothetical prote
47	43.5	39.9	354	2	T12741	hypothetical prote
48	43.5	39.9	801	1	S00943	glucose dehydrogen
49	43	39.4	266	2	S51052	glutamate CoA-tr
50	43	39.4	281	2	H71308	hypothetical prote
51	43	39.4	318	2	A84466	hypothetical prote
52	43	39.4	352	2	G69774	hypothetical prote
53	43	39.4	478	2	S33886	DNA-directed RNA p
54	43	39.4	497	2	T04731	cytochrome P450 ho
55	43	39.4	500	2	C85441	cytochrome P450-11
56	43	39.4	500	2	T52174	aconitate hydratase
57	43	39.4	780	2	T52543	hypothetical prote
58	43	39.4	963	2	T19140	spectrin beta chain
59	43	39.4	2137	1	SJHUB	probable pre-tRNA
60	42.5	39.0	978	2	T40803	hypothetical prote
61	42	38.5	176	2	A86441	assemblin, striate
62	42	38.5	276	2	T08178	hypothetical 45.1
63	42	38.5	408	2	H65137	omega-3 fatty acid
64	42	38.5	438	2	T15039	CAMP-dependent pro
65	42	38.5	464	2	A32461	complement C9 prec
66	42	38.5	559	1	C9HU	2-aminobenzoate-Co
67	42	38.5	603	2	S22402	hypothetical prote
68	42	38.5	638	2	T13691	hypothetical prote
69	42	38.5	679	2	S48437	toxin secretion intia
70	42	38.5	704	2	H82381	translation intia
71	42	38.5	918	2	T38786	fumylate reductase
72	41.5	38.1	230	2	H70349	polyketide synthase
73	41	37.6	113	2	JC5857	translation elonga
74	41	37.6	396	1	A54536	anaerobic glycerol
75	41	37.6	415	2	AH0465	probable membrane
76	41	37.6	432	2	D86937	hypothetical prote
77	41	37.6	434	2	S37907	probable cytochrom
78	41	37.6	481	2	B96691	probable proline o
79	41	37.6	492	2	T41543	probable FAD bindi
80	41	37.6	571	2	T39378	hypothetical prote
81	41	37.6	607	2	T20796	hypothetical prote
82	41	37.6	614	2	T20795	hypothetical prote
83	41	37.6	654	2	AD3183	peptidase [importe
84	41	37.6	691	2	S48330	hypothetical prote
85	41	37.6	709	2	T00664	hypothetical prote
86	41	37.6	1021	2	G01202	NaCl electroneutra
87	41	37.6	1780	2	T17272	hypothetical prote
88	41	37.6	1805	2	A34736	nestin - rat
89	41	37.6	1882	1	GNVVR	genome polypotein
90	41	37.6	6642	2	T29757	protein UNC-89 C
91	40.5	37.2	359	2	T33115	hypothetical prote
92	40.5	37.2	470	2	D90895	probable aldehyde
93	40.5	37.2	470	2	D85722	probable aldehyde
94	40	36.7	138	2	T43419	ribosomal protein
95	40	36.7	140	2	T39778	40s ribosomal prot
96	40	36.7	198	2	D88098	protein f06b4.5 [i
97	40	36.7	198	2	T32025	hypothetical prote
98	40	36.7	205	2	JC7975	HIV accessory prot
99	40	36.7	219	2	AD2776	transcription regu
100	40	36.7	227	2	S44836	K2D10.4 protein -





```

A:Accession: PQ0429
A:Molecule type: mRNA
A:Residues: 1-56 <ROB>
A:Note: this protein is from the VP1/2A Junction region
C:Superfamily: hepatitis A virus genome polypeptide
C:Keywords: coat protein; core protein; polypeptide

          97.2%; Score 106; DB 2; Length 56;
Best Local Similarity 95.0%; Pred. No. 1.5e-09;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1  VDDPRSEEDKRFESHIECKR 40
          |||||:|||||
Db      16  VDDPRSEEDKRFESHIECKR 35

RESULT 9
S04137
genome polypeptide - human hepatitis A virus (strain LDC-1) (fragment)
C:Species: human hepatitis A virus
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 21-Jul-2000
C:Accession: S04137
F:Andonov, A.P.; Lau, P.C.K.; Chaudhary, R.
Nucleic Acids Res. 17, 3594, 1989
A:Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A virus
A:Reference number: S04137; PMID:89263805; PMID:2542903

```

```

Qy      1 VDDPRSEEDKPFESHCCK 20
        |||||:|||||
Db      283 VDDPRSEEDRPFESHCCK 302

```

```

Qy      1 VDDPRSEEDKPFESHCCK 20
        |||||:|||||
Db      283 VDDPRSEEDRPFESHCCK 302

```



Db 16 VDDPRSEEDKRFESHIEYRK 35  
|||||:|||||  
Query Match 77.1%; Score 84; DB 2; Length 56;  
Best Local Similarity 94.1%; Pred. No. 3.8e-06;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 15  
PQ0434  
genome polyprotein - human hepatitis A virus (strain KPH) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0434  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A>Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0434  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A>Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 83.5%; Score 91; DB 2; Length 56;  
Best Local Similarity 85.0%; Pred. No. 3.1e-07;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECKK 20  
|||||:|||||  
Db 16 VDDPRSEEDKRFESHIEYRK 35  
|||||:|||||

RESULT 16  
PQ0437  
genome polyprotein - human hepatitis A virus (strain JM55) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0437  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A>Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0437  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A>Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 77.1%; Score 84; DB 2; Length 56;  
Best Local Similarity 94.1%; Pred. No. 3.8e-06;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIE 17  
|||||:|||||  
Db 16 VDDPRSAEDKRFESHIE 32  
|||||:|||||

RESULT 17  
PQ0436  
genome polyprotein - human hepatitis A virus (strain AGM27) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0436  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A>Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0436  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A>Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 77.1%; Score 84; DB 2; Length 56;  
Best Local Similarity 94.1%; Pred. No. 3.8e-06;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIE 17  
|||||:|||||  
Db 16 VDDPRSAEDKRFESHIE 32  
|||||:|||||

RESULT 18  
GNNYSA  
genome polyprotein - simian hepatitis A virus (strain AGM-27)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro  
C:Species: simian hepatitis A virus  
C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jun-2000  
C:Accession: A30470; S04885; S03965  
R:Tsarev, S.A.  
submitted to JIPID, April 1991  
A:Reference number: A30470  
A:Accession: A30470  
A:Molecule type: Genomic RNA  
A:Residues: 1-2230 <TSA>  
A:Cross-references: GB:D00924; NID:G222597; PIDN:BA00766.1; PID:G222598  
R:Tsarev, S.A.; Emerson, S.U.; Balayan, M.S.; Ticehurst, J.; Purcell, R.H.  
J. Gen. Virol. 72, 1677-1683, 1991  
A>Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure an  
A:Reference number: JQ1080; MUID:91311420; PMID:1649901  
A:Contents: annotation  
A>Note: neither amino acid nor nucleotide sequence is given  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhiko  
submitted to the EMBL Data Library, May 1989  
A:Reference number: S04885  
A:Accession: S04885  
A:Molecule type: Genomic RNA  
A:Residues: 1750-2164 <BAL>  
A:Cross-references: EMBL:X15461; NID:G61971; PIDN:CAA33490.1; PID:G930268  
R:Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhiko  
FEBS Lett. 247, 425-428, 1989  
A>Title: Variations in genome fragments coding for RNA polymerase in human and simian he  
A:Reference number: S03965; MUID:89232168; PMID:2541023  
A:Accession: S03965  
A:Molecule type: Genomic RNA  
A:Residues: 1960-2164 <BAL2>  
A:Cross-references: EMBL:X15461  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-27/Product: coat protein 1A #status predicted <C1A>  
F:28-249/Product: coat protein 1B #status predicted <C1B>  
F:250-495/Product: coat protein 1C #status predicted <C1C>  
F:496-795/Product: coat protein 1D #status predicted <C1D>  
F:796-984/Product: core protein 2A #status predicted <C2A>  
F:985-1091/Product: core protein 2B #status predicted <C2B>  
F:1092-1426/Product: core protein 2C #status predicted <C2C>  
F:1427-1498/Product: protein 3A #status predicted <P3A>  
F:1499-1521/Product: protein 3B #status predicted <P3B>  
F:1522-1741/Product: protein 3C #status predicted <P3C>  
F:1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 77.1%; Score 84; DB 1; Length 2230;  
Best Local Similarity 94.1%; Pred. No. 0.00019;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIE 17  
|||||:|||||  
Db 783 VDDPRSAEDKRFESHIE 799  
|||||:|||||

RESULT 19  
PQ0433  
genome polyprotein - human hepatitis A virus (strain PA21) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0433

R;Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel, J. Gen. Virol. 73, 1365-1377, 1992

A;Title: Genetic relatedness of hepatitis A virus strains recovered from different geographical regions

A;Reference number: PQ0427; MUID:92300330; PMID:1318940

A;Accession: PQ0433

A;Molecule type: mRNA

A;Residues: 1-55 <ROB>

A;Note: this protein is from the VP1/2A Junction region

C;Superfamily: hepatitis A virus genome polyprotein

C;Keywords: coat protein; core protein; polyprotein

Query Match 74.3%; Score 81; DB 2; Length 55;  
Best Local Similarity 82.4%; Pred. No. 1.1e-05; Mismatches 0; Indels 0; Gaps 0;  
Matches 14; Conservative 3;

QY 1 VDDPRSEEDKRFESHIE 17  
|||||:|||||  
DB 15 VDDPRTEEDKRFESHIE 31

RESULT 20

PQ0435

genome polyprotein - human hepatitis A virus (strain Cy145) (fragment)

C;Species: human hepatitis A virus

C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999

C;Accession: PQ0435

R;Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel, J. Gen. Virol. 73, 1365-1377, 1992

A;Title: Genetic relatedness of hepatitis A virus strains recovered from different geographical regions

A;Reference number: PQ0427; MUID:92300330; PMID:1318940

A;Accession: PQ0435

A;Molecule type: mRNA

A;Residues: 1-55 <ROB>

A;Note: this protein is from the VP1/2A Junction region

C;Superfamily: hepatitis A virus genome polyprotein

C;Keywords: coat protein; core protein; polyprotein

Query Match 74.3%; Score 81; DB 2; Length 55;  
Best Local Similarity 82.4%; Pred. No. 1.1e-05; Mismatches 0; Indels 0; Gaps 0;  
Matches 14; Conservative 3;

QY 1 VDDPRSEEDKRFESHIE 17  
|||||:|||||  
DB 16 VDDPRTEEDKRFESHIE 32

RESULT 21

GNYS2

genome polyprotein - simian hepatitis A virus (strain CY-145) (fragment)

N;Contains: amino end of core protein 2A; coat protein 1A; coat protein 1B; coat protein 1C

C;Species: simian hepatitis A virus

A;Note: host Macaca fascicularis (cynomolgus macaque)

C;Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 16-Jul-1999

C;Accession: JQ1180

R;Nainan, O.V.; Margolis, H.S.; Robertson, B.H.; Balayan, M.; Brinton, M.A.; J. Gen. Virol. 72, 1685-1689, 1991

A;Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus macaque

A;Reference number: JQ1180; MUID:913111421; PMID:1649902

A;Accession: JQ1180

A;Molecule type: genomic RNA

A;Residues: 1-839 <NAI>

A;Cross-references: GB:M59286; NID:G329599; PID:AAA45473.1; PID:g55083

C;Superfamily: hepatitis A virus genome polyprotein

C;Keywords: coat protein; core protein; glycoprotein; polyprotein

F;1-23/Product: coat protein 1A #status predicted <VP0>

F;24-245/Product: coat protein 1B #status predicted <VP3>

F;246-491/Product: coat protein 1C #status predicted <VP1>

F;492-839/Product: core protein 2A (fragment) #status predicted <P2P>

F;261,312,728,756/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 74.3%; Score 81; DB 1; Length 839;  
Best Local Similarity 82.4%; Pred. No. 0.00019; Mismatches 3; Indels 0; Gaps 0;  
Matches 14; Conservative 3;

A;Note: amino end of a C3-derived peptide designated exudate neutrophil chemotactic factor  
A;Accession: P25551  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-754 <JOH>  
A;Cross-references: EMBL:U80847; PIDN:AA37983.1; GSPDB:GN00028; CESP:C17H11.2  
A;Experimental source: strain Bristol N2; clone C17H11  
C;Genetics:  
A;Gene: CESP:C17H11.2  
A;Map position: X  
A;Introns: 155/1; 246/3; 288/3; 368/3; 631/1  
C;Superfamily: Caenorhabditis elegans hypothetical protein C17H11.2

Query Match 46.3%; Score 50.5; DB 2; Length 754;  
Best Local Similarity 37.5%; Pred. No. 9.2;  
Matches 9; Conservative 7; Mismatches 3; Indels 5; Gaps 1;

QY 2 DDPSEEDK----RFESHIECR 20  
DB 580 DEPLTEQRIHKRFYVYKCR 603

RESULT 25  
C3RT  
complement C3 precursor - rat  
N;Alternate names: 37K phospholipase A2 inhibitory protein  
N;Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subunit;  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 04-Dec-1992 #sequence revision 07-Oct-1994 #text change 18-Jun-1999  
C;Accession: S15764; A54562; A01260; B35979; A35979; PNO566; A32281; S08592  
R;Mitsumi, Y.; Kohda, M.; Ikehara, Y.  
Nucleic Acids Res. 18, 2178, 1990  
A;Title: Nucleotide and deduced amino acid sequence of rat complement C3.  
A;Reference number: S15764; MUID:90245672; PMID:2336397  
A;Accession: S15764  
A;Molecule type: mRNA  
A;Residues: 1-1663 <MIS>  
A;Cross-references: EMBL:X52477; NID:g56953; PIDN:CAA36716.1; PID:g56954  
R;Sundstrom, S.A.; Komm, B.S.; Ponce-de-Leon, H.; Yi, Z.; Teuscher, C.; Lyttle, C.R.  
J. Biol. Chem. 264, 16941-16947, 1999  
A;Title: Estrogen regulation of tissue-specific expression of complement C3.  
A;Reference number: A54562; MUID:189380332; PMID:2674144  
A;Accession: A54562  
A;Status: translation not shown  
A;Molecule type: mRNA  
A;Residues: P',1316-1595 <SUN>  
A;Cross-references: GB:M29866; NID:Q203200; PIDN:AAA40837.1; PID:g554423  
R;Jacobs, J.W.; Rubin, J.S.; Huggli, T.E.; Bogardt, R.A.; Mariz, I.K.; Daniels, J.S.; Dau  
Biochemistry 17, 5031-5038, 1978  
A;Title: Purification, characterization, and amino acid sequence of rat anaphylatoxin (C  
A;Reference number: A01260; MUID:79062262; PMID:309768  
A;Accession: A01260  
A;Molecule type: protein  
A;Residues: 671-703,'K',705-720,'KL',723-748 <JAC>  
A;Note: three disulfide bonds are present  
R;Suwa, Y.; Kudo, I.; Imaizumi, A.; Okada, M.; Kamimura, T.; Suzuki, Y.; Chang, H.W.; Ha  
Proc. Natl. Acad. Sci. U.S.A. 87, 2395-2399, 1990  
A;Title: Proteinaceous inhibitors of phospholipase A-2 purified from inflammatory sites  
A;Reference number: A35979; MUID:90207203; PMID:2320562  
A;Accession: B35979  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: X',998-1005 <SUW>  
A;Accession: A35979  
A;Molecule type: protein  
A;Residues: X',961-962,'P',964-969 <SUZ>  
R;Kakagawa, H.; Komorita, N.  
Biochem. Biophys. Res. Commun. 194, 1181-1187, 1993  
A;Title: Complement component C3-derived neutrophil chemotactic factors purified from ex  
A;Reference number: PNO566; MUID:93356786; PMID:8352775  
A;Accession: PNO567  
A;Molecule type: protein  
A;Residues: 568-592 <NAK>

A;Note: amino end of a C3-derived peptide designated exudate neutrophil chemotactic factor  
A;Accession: PNO566  
A;Molecule type: protein  
A;Residues: 671-687 <NA2>  
A;Note: amino end of peptide designated neutrophil chemotactic factor 1 and probably ide  
R;Kuivanen, P.C.; Capulong, R.B.; Harkins, R.N.; DeSombre, E.R.  
Biochem. Biophys. Res. Commun. 158, 898-905, 1989  
A;Title: The estrogen-responsive 110K and 74K rat uterine secretory proteins are structur  
A;Reference number: A32281; MUID:89149812; PMID:2645873  
A;Accession: A32281  
A;Molecule type: protein  
A;Residues: 25-41 <KUI>  
A;Experimental source: 17beta-estradiol-stimulated uterus of immature rat  
A;Note: the authors treat this 74K uterine secretory protein, identical as far as sequen  
ent  
C;Comment: Complement C3 contains two chains, formed by removal of four residues and lin  
alternative complement pathways, releases the C3a anaphylatoxin from the amino end of t  
native-complement-pathway C3/C5 convertase.  
C;Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.  
C;Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign pa  
e classical-complement-pathway C3/C5 convertase. The activity of C3b is regulated by prot  
C;Comment: The major site of synthesis of this plasma protein is the liver.  
C;Superfamily: alpha-2-macroglobulin  
C;Keywords: acute phase; chemotaxis; complement alternate pathway; complement pathway; G;  
F;1-24/Domain: signal sequence #status predicted <SIG>  
F;25-666/Product: complement C3 and C3b beta chain #status predicted <C3BB>  
F;25-666,671-1663/Product: complement C3 #status predicted <CC3>  
F;25-666,749-1663/Product: complement C3b #status predicted <C3B>  
F;671-1663/Product: complement C3 alpha chain #status predicted <CC3A>  
F;671-748/Product: C3a anaphylatoxin #status experimental <C3T>  
F;749-1663/Product: complement C3b alpha' chain #status predicted <C3BA>  
F;946-1303/Product: C3dk fragment #status predicted <CDK>  
F;1002-1303/Product: C3d fragment #status predicted <C3D>  
F;1424-1457/Region: properdin binding  
F;558-616,626-661,693-720,694-727,707-728,873-1513,1101-1158,1358-1489,1389-1458,1506-151  
F;748-749/Cleavage site: Arg-Ser (C3 convertase) #status predicted  
F;939,1617/Binding site: carbonyl (Asn) (covalent) #status predicted  
F;1010-1013/Cross-link: thiolster (Cys-Gln) #status predicted  
F;1303-1304/Cleavage site: Arg-Ser (complement factor I) #status predicted  
F;1320-1321/Cleavage site: Arg-Ser (complement factor I) #status predicted

Query Match 45.9%; Score 50; DB 1; Length 1663;  
Best Local Similarity 38.9%; Pred. No. 25;  
Matches 7; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 2 DDPSEEDKRFESHIECR 19  
DB 1574 DEVQAGQERFISHVKCR 1591

RESULT 26  
C3GP  
complement C3 precursor - guinea pig  
N;Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subunit; C  
C;Species: Cavia porcellus (guinea pig)  
C;Date: 07-Feb-1992 #sequence revision 07-Oct-1994 #text change 18-Jun-1999  
C;Accession: A37156; S03375; A20342; D20342; D20342; A31222  
R;Auerbach, H.S.; Burger, R.; Dodds, A.; Colten, H.R.  
J. Clin. Invest. 86, 96-106, 1990  
A;Title: Molecular basis of complement C3 deficiency in guinea pigs.  
A;Reference number: A37156; MUID:90307998; PMID:1973176  
A;Accession: A37156  
A;Molecule type: mRNA  
A;Residues: 1-1666 <AUE>  
A;Cross-references: GB:M34054; NID:g191262; PIDN:AAA37038.1; PID:g305335  
R;Gerard, N.P.; Lively, M.O.; Gerard, C.  
Protein Seq. Data Anal. 1, 473-478, 1988  
A;Title: Amino acid sequence of guinea pig C3a anaphylatoxin.  
A;Reference number: S03375; MUID:89113342; PMID:3064079  
A;Accession: S03375  
A;Molecule type: protein  
A;Residues: 676-730,'N',732-752 <GER>  
A;Experimental source: complement-activated guinea pig serum

RESULT 27  
A33229  
complement C3 - African clawed frog (fragment)  
NT:Contains: alternative-complement-pathway C3/C5 convertase (SC 3.4.21.47) C3b subunit;  
Species: Xenopus laevis (African clawed frog)  
CDate: 31-Jul-1989 #sequence\_revision 31-Jul-1989 #text\_change 07-Oct-1994  
CAccession: A33229  
R:Grossberger, D.; Marcuz, A.; Du Pasquier, L.; Lambris, J.D.  
Proc. Natl. Acad. Sci. U.S.A. 86, 1323-1327, 1989  
A:Title: Conservation of structural and functional domains in complement component C3 of  
A:Reference number: A33229; MUID:89145234; PMID:2919181  
A:Accession: A33229  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-322 <GRO>  
A:Cross-references: GB:M21790; GB:J04493  
C:Comment: Complement C3 contains two chains, formed by removal of four residues and lin  
n alternative complement pathways, releases the C3a anaphylatoxin from the amino end of t  
n alternative-complement-pathway C3/C5 convertase.  
C:Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.

```

A:Accession: B84699
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-454 <STO>
A:Cross-references: GB:AE002093; NID:g3582340; PIDN:AAC35237.1; GSPDB:GN00139
A:Gene: At2g29670
A:Map position: 2

Query Match      42.7%; Score 46.5; DB 2; Length 454;
Best local similarity 50.0%; Pred. No. 22;
Matches 10; Conservative 6; Mismatches 1; Gaps 1;

          1 VDDPRSEED-KRFESHIECR 19
            :|:|:|:|:|:|:|:|:|
389 IDDSLQDETGRKFVSHVEAR 408

```

```
RESULT 30
E83077
conserved hypothetical protein PA4543 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83077
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: E83077
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-242 <STO>
A:Cross-references: GB:AE004868; GB:AE004091; NID:g9950785; PIDN:AAG07931.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4543
C:Superfamily: Streptomyces coelicolor hypothetical protein SC4A10.14c

Query Match 42.2%; Score 46; DB 2; Length 242;
Best Local Similarity 45.0%; Pred. No. 14;
Matches 9; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

QY 2 DDPRSEED--KRFESHIECR 19
Db 42 DDPRAVEENRRLLTERLECR 61

Search completed: May 11, 2004, 13:41:14
Job time : 12.75 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: May 11, 2004, 13:24:26 ; Search time 6.75 Seconds  
(without alignments)  
154.282 Million cell updates/sec

Title: US-09-171-432a-38

Perfect score: 109

Sequence: 1 VDDFRSEEDKRFESHIECRK 20

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	109	100.0	2226	1	POLG_HPAV2
2	109	100.0	2226	1	POLG_HPAV4
3	109	100.0	2226	1	POLG_HPAV8
4	109	100.0	2227	1	POLG_HPAVH
5	109	100.0	2227	1	POLG_HPAVM
6	106	97.2	341	1	POLG_HPAV1
7	106	97.2	852	1	POLG_HPAVC
8	106	97.2	2227	1	POLG_HPAV3
9	88	80.7	808	1	POLG_HPAV6
10	84	77.1	2230	1	POLG_HPAVS
11	81	74.3	839	1	POLG_HPAVT
12	50	45.9	1663	1	CO3_RAT
13	50	45.9	1666	1	CO3_CAVPO
14	48	44.0	323	1	CO3_XENLA
15	46	42.2	242	1	Y983_PSEAE
16	46	42.2	1663	1	CO3_MOUSE
17	45	41.3	741	1	YD35_MYCPN
18	44	40.4	222	1	RL4_CHLTR
19	44	40.4	276	1	TRME_HUMAN
20	44	40.4	1101	1	FURC_DROME
21	43.5	39.9	801	1	DHGA_ACICA
22	43	39.4	265	1	GCTB_ACIFE
23	43	39.4	267	1	SURE_METWA
24	43	39.4	478	1	RPB1_EUPOC
25	43	39.4	502	1	C8LD_ARATH
26	43	39.4	507	1	DNAL_FUSNN
27	43	39.4	780	1	ACON_HUMAN
28	43	39.4	963	1	YQ36_CABEL
29	43	39.4	2137	1	SPCB_HUMAN
30	42	38.5	269	1	SPAS_CHLEU
31	42	38.5	276	1	SPAS_CHLEU
32	42	38.5	408	1	RTCB_ECOLI
33	42	38.5	464	1	AK15_RAT

34	42	38.5	559	1	CO9_HUMAN
35	42	38.5	639	1	DNAL_SHEON
36	42	38.5	679	1	YIS3_YEAST
37	42	38.5	918	1	IF38_SCHPO
38	41	37.6	267	1	SURE_METAC
39	41	37.6	396	1	EFTU_FLESI
40	41	37.6	434	1	YK12_YEAST
41	41	37.6	492	1	PROD_SCHPO
42	41	37.6	571	1	YB63_SCHPO
43	41	37.6	691	1	YIO4_YEAST
44	41	37.6	925	1	NRP2_RAT
45	41	37.6	931	1	NRP2_HUMAN
46	41	37.6	931	1	NRP2_MOUSE
47	41	37.6	1021	1	S123_HUMAN
48	41	37.6	1805	1	NEST_RAT
49	41	37.6	1882	1	POL2_TRSVR
50	41	37.6	6632	1	UN89_CABEL
51	40	36.7	140	1	RS16_SCHPO
52	40	36.7	227	1	YMQ4_CABEL
53	40	36.7	275	1	YCT7_YEAST
54	40	36.7	412	1	EP1H_YEAST
55	40	36.7	435	1	RP0D_SORBI
56	40	36.7	466	1	MM13_RAT
57	40	36.7	747	1	VIVC_BPT7
58	40	36.7	1355	1	VG46_HSV11
59	40	36.7	1527	1	RP0D_MAIZE
60	40	36.7	3432	1	POLG_JAEV1
61	40	36.7	3432	1	POLG_JAEVS
62	40	36.7	3432	1	POT2_STYPL
63	39	35.8	127	1	ATF1_RAT
64	39	35.8	147	1	KCY_MOUSE
65	39	35.8	196	1	RLA_CHLMU
66	39	35.8	222	1	ATE_RHIME
67	39	35.8	259	1	TRMB_MOUSE
68	39	35.8	268	1	SPAS_DUNBI
69	39	35.8	277	1	OSTP_SHEEP
70	39	35.8	310	1	DHLA_XANAU
71	39	35.8	353	1	ABR2_MOUSE
72	39	35.8	401	1	PGK_HALVA
73	39	35.8	419	1	ENO_PYRAE
74	39	35.8	482	1	MURE_THETN
75	39	35.8	522	1	RF3_LISIN
76	39	35.8	522	1	RF3_LISMO
77	39	35.8	522	1	SBP_SOYBN
78	39	35.8	524	1	SR72_SCHPO
79	39	35.8	561	1	TDR4_HUMAN
80	39	35.8	633	1	AMO_PEA
81	39	35.8	674	1	TRDN_HUMAN
82	39	35.8	728	1	DAC1_XENLA
83	39	35.8	824	1	CHSC_ASPTU
84	39	35.8	889	1	PHSL_VICFA
85	39	35.8	1003	1	PTPX_MACNE
86	39	35.8	1013	1	PTK7_CHICK
87	39	35.8	1051	1	THY7_PIG
88	39	35.8	1242	1	DHVS_DIACA
89	38.5	35.3	373	1	GLYA_BUCAI
90	38.5	35.3	417	1	YCA4_HUMAN
91	38.5	35.3	733	1	MYH7_PIG
92	38.5	35.3	1935	1	ANK2_HUMAN
93	38.5	35.3	3924	1	YNP4_CABEL
94	38	34.9	78	1	YPT2_CABEL
95	38	34.9	189	1	CSE4_YEAST
96	38	34.9	229	1	DHXS_MOUSE
97	38	34.9	280	1	OSTP_MOUSE
98	38	34.9	294	1	OSTP_PIG
99	38	34.9	303	1	RLA0_PLAF8
100	38	34.9	316	1	RLA0_PLAF8

## ALIGNMENTS

RESULT 1



POLG HPAV2 STANDARD; PRT; 2226 AA.

AC P26580;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)]

DE P3D (EC 2.7.7.48)

OS Hepatitis A virus (strain 24a)

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12094;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=91162758; PubMed=1705995;

RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,

RA "Antigenic and genetic variation in cytopathic hepatitis A virus

RT variants arising during persistent infection: evidence for genetic

RT recombination."

RL J. Virol. 65:2056-2065 (1991).

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +

CC [RNA] (N).

CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,

CC each of which is composed of one copy each of proteins VP1, VP2,

CC VP3, and VP4.

CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; M59810; AAA45468.1; -.

DR MEROPS; C03.005; -.

DR InterPro; IPR004004; Calici\_pol\_hel.

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR000605; RNA\_helicase.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.

DR InterPro; IPR001205; RNA\_pol\_P3D.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR InterPro; IPR008975; Viral\_cap\_coat.

DR Pfam; PF00680; RNA\_dep\_RNA\_pol\_1.

DR Pfam; PF00910; RNA\_helicase\_1.

DR PRINTS; PR00918; CALICIVIRUSN.

DR Polyprotein; Coat protein; Core protein; Transferase;

KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.

KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.

FT CHAIN 1 23

FT CHAIN 24 245

FT CHAIN 246 491

FT CHAIN 492 794

FT CHAIN 795 900

FT CHAIN 901 1087

FT CHAIN 1088 1422

FT CHAIN 1423 1495

FT CHAIN 1496 1518

FT CHAIN 1519 1737

FT CHAIN 1738 2226

SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;

Query Match 100.0%; Score 109; DB 1; Length 2226;

Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKK 20

DB 779 VDDPRSEEDKRFESHIECKK 798

RESULT 2

POLG HPAV4 STANDARD; PRT; 2226 AA.

AC P26581;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)]

DE P3D (EC 2.7.7.48)

OS Hepatitis A virus (strain 43c)

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12095;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=91162758; PubMed=1705995;

RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,

RA "Antigenic and genetic variation in cytopathic hepatitis A virus

RT variants arising during persistent infection: evidence for genetic

RT recombination."

RL J. Virol. 65:2056-2065 (1991).

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +

CC [RNA] (N).

CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,

CC each of which is composed of one copy each of proteins VP1, VP2,

CC VP3, and VP4.

CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; M59809; AAA45469.1; -.

DR MEROPS; C03.005; -.

DR InterPro; IPR004004; Calici\_pol\_hel.

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR000605; RNA\_helicase.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.

DR InterPro; IPR001205; RNA\_pol\_P3D.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR InterPro; IPR008975; Viral\_cap\_coat.

DR Pfam; PF00680; RNA\_dep\_RNA\_pol\_1.

DR Pfam; PF00910; RNA\_helicase\_1.

DR PRINTS; PR00918; CALICIVIRUSN.

DR Polyprotein; Coat protein; Core protein; Transferase;

KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.

KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.

FT CHAIN 1 23

FT CHAIN 24 245

FT CHAIN 246 491

FT CHAIN 492 794

FT CHAIN 795 900

FT CHAIN 901 1087

FT CHAIN 1088 1422

FT CHAIN 1423 1495

FT CHAIN 1496 1518

FT CHAIN 1519 1737

FT CHAIN 1738 2226

SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B9BF75 CRC64;

Query Match 100.0%; Score 109; DB 1; Length 2226;

Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKK 20

```
Db 779 VDDPRSEEDKRFESHIECRK 798
RESULT 3
POLG_HPAV8 STANDARD; PRT; 2226 AA.
AC P26582;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain 18f).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91162758; PubMed=1705995;
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,
RA "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination."
RL J. Virol. 65:2056-2065 (1991).
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
CC each of which is composed of one copy each of proteins VP1, VP2,
CC VP3, and VP4.
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M59808; AAA45467.1; -.
CC PDB; 1QA7; 15-MAY-00.
CC
CC MEROPS; C03.005; -.
CC InterPro; IPR004004; Calici pol hel.
CC InterPro; IPR009003; Cys Ser triypsin.
CC InterPro; IPR008605; RNA helicase.
CC InterPro; IPR007095; RNA pol PS.
CC InterPro; IPR001205; RNA pol P3D.
CC InterPro; IPR007094; RNA pol psvir.
CC InterPro; IPR008975; Viral cap coat.
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Pfam; PF00910; RNA_helicase; 1.
CC PRINTS; PR00918; CALICVIRUSNS.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.
FT CHAIN 1 23
FT CHAIN 24 245
FT CHAIN 246 491
FT CHAIN 492 794
FT CHAIN 795 900
FT CHAIN 901 1087
FT CHAIN 1088 1422
FT CHAIN 1423 1495
FT CHAIN 1496 1518
FT CHAIN 1519 1737
FT CHAIN 1738 2226
FT CHAIN 2226 AA; 251292 MW; 2496A63396C8D6B CRC64;
SQ SEQUENCE 2226 AA; 251292 MW; 2496A63396C8D6B CRC64;

Query Match 100.0%; Score 109; DB 1; Length 2226;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20
DB 779 VDDPRSEEDKRFESHIECRK 798
RESULT 4
POLG_HPAVH STANDARD; PRT; 2227 AA.
AC P08617; P06443; Q81082;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
OS Hepatitis A virus (strain HM-175).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12098;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87061253; PubMed=3023706;
RA Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A.,
RA Baroudy B.M.;
RA "Complete nucleotide sequence of wild-type hepatitis A virus:
RT comparison with different strains of hepatitis A virus and other
RT picornaviruses."
RL J. Virol. 61:50-59 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87175701; PubMed=3031686;
RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstone S.M.,
RA Purcell R.H.;
RA "Complete nucleotide sequence of an attenuated hepatitis A virus:
RT comparison with wild-type virus."
RL Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501 (1987).
RN [3]
RP SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.
RX MEDLINE=85166289; PubMed=2984684;
RA Baroudy B.M., Ticehurst J.R., Miele T.A., Maizel J.V. Jr.,
RA Purcell R.H., Feinstone S.M.;
RA "Sequence analysis of hepatitis A virus cDNA coding for capsid
RT proteins and RNA polymerase."
RL Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147 (1985).
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
CC each of which is composed of one copy each of proteins VP1, VP2,
CC VP3, and VP4.
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -!- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED
CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -!- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT
CC SHOWN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14114; AAA45475.1; -.
CC EMBL; M14707; AAA45465.1; -.
CC EMBL; M14707; AAA45466.1; ALT_INIT.
CC EMBL; M16632; AAA45471.1; -.
CC PIR; A03905; A03905.
CC PIR; A25981; GNNYHW.
```

DR PIR; A94149; GNNYMK.  
 DR PDB; 1HAV; 23-DEC-96.  
 DR MEROPS; C03.005; --  
 DR InterPro; IPR004004; Calici\_pol\_hel.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR000605; RNA\_helicase.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR001205; RNA\_pol\_P3D.  
 DR InterPro; IPR007094; RNA\_pol\_P5vir.  
 DR InterPro; IPR008975; Viral\_cap\_coat.  
 DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR Pfam; PF00910; RNA\_helicase; 1.  
 DR PRINTS; PR00918; CALICIVIRUSNS.  
 KW Polyprotein; Coat protein; Core protein; Transferase;  
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.  
 FT CHAIN 1 23  
 FT CHAIN 24 245  
 FT CHAIN 246 491  
 FT CHAIN 492 836  
 FT CHAIN 837 980  
 FT CHAIN 981 1087  
 FT CHAIN 1088 1422  
 FT CHAIN 1423 1496  
 FT CHAIN 1497 1519  
 FT CHAIN 1520 1738  
 FT CHAIN 1739 2227  
 FT CHAIN 77 77  
 FT CHAIN 764 764  
 FT CHAIN 821 821  
 FT CHAIN 1052 1052  
 FT CHAIN 1062 1062  
 FT CHAIN 1118 1118  
 FT CHAIN 1151 1151  
 FT CHAIN 1163 1163  
 FT CHAIN 1277 1277  
 FT CHAIN 1500 1500  
 FT CHAIN 1805 1805  
 FT CHAIN 1930 1930  
 SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AEB740A6 CRC64;

Query Match 100.0%; Score 109; DB 1; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECRK 20  
 |||||  
 Db 779 VDDPRSEEDKRFESHIECRK 798

RESULT 5  
 POLG\_HPAVM STANDARD; PRT; 2227 AA.  
 AC P13901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;  
 AC Q81090; Q81091; Q81092; Q81093;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
 DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
 DE P3D (EC 2.7.7.48)].  
 OS Hepatitis A virus (strain MBB).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=12100;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88045071; PubMed=2823500;  
 RA Paul A.V., Tada H., der Helm K., Wessel T., Kiehn T., Wimmer E.,  
 RA Deinhardt F.;  
 RA "The entire nucleotide sequence of the genome of human hepatitis A  
 RT virus (isolate MBB).";  
 RL Virus Res. 8:153-171(1987).  
 RL NUCLEOTIDE ACTIVITY: N nucleoside triphosphate = N diphosphate +

CC {RNA} (N).  
 CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -!- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT  
 CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M20273; AAA5474.1; --  
 CC MEROPS; C03.005; --  
 DR InterPro; IPR004004; Calici\_pol\_hel.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR000605; RNA\_helicase.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR001205; RNA\_pol\_P3D.  
 DR InterPro; IPR007094; RNA\_pol\_P5vir.  
 DR InterPro; IPR008975; Viral\_cap\_coat.  
 DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR Pfam; PF00910; RNA\_helicase; 1.  
 DR PRINTS; PR00918; CALICIVIRUSNS.  
 KW Polyprotein; Coat protein; Core protein; Transferase;  
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
 FT CHAIN 1 23  
 FT CHAIN 24 245  
 FT CHAIN 246 491  
 FT CHAIN 492 836  
 FT CHAIN 837 980  
 FT CHAIN 981 1087  
 FT CHAIN 1088 1422  
 FT CHAIN 1423 1496  
 FT CHAIN 1497 1519  
 FT CHAIN 1520 1738  
 FT CHAIN 1739 2227  
 SQ SEQUENCE 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;

Query Match 100.0%; Score 109; DB 1; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECRK 20  
 |||||  
 Db 779 VDDPRSEEDKRFESHIECRK 798

RESULT 6  
 POLG\_HPAVI STANDARD; PRT; 341 AA.  
 ID P13672;  
 AC P13672;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Coat protein VP3 (1C); Coat protein VP1  
 DE (1D); Core protein P2A (Fragment).  
 OS Hepatitis A virus (strain LCDC-1).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=12093;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89263805; PubMed=2542903;  
 RA Andonov A.P., Lau P., Chaudhary R.;  
 RA "Nucleotide sequence of the VP1 gene from a Chinese strain of  
 RT hepatitis A virus (HAV).";  
 RL Nucleic Acids Res. 17:3594-3594(1989).

```
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units.  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; X14666; CAA32794.1; -.  
DR PIR; S04137; S04137.  
DR InterPro; IPR008975; Viral_cap_coat.  
KW Polyprotein; Coat protein; Core protein.  
FT NON_TER 1 1  
FT CHAIN <1 1 COAT PROTEIN VP3.  
FT CHAIN 2 340 COAT PROTEIN VP1.  
FT CHAIN 341 >341 CORE PROTEIN P2A.  
FT NON_TER 341 341  
SQ SEQUENCE 341 AA; 38003 MW; 066918289BF126D5 CRC64;  
  
Query Match 97.2%; Score 106; DB 1; Length 341;  
Best Local Similarity 95.0%; Pred. No. 6.4e-09;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 263 VDDPRSEEDKRFESHIECRK 302  
[1]  
-----  
RESULT 7  
POLG_HPAVC STANDARD; PRT; 852 AA.  
AC P06442; Q83741; Q83742;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein  
DE P2A] (Fragment).  
OS Hepatitis A virus (strain CR326).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI_TaxID=12097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85185648; PubMed=2985793;  
RA Linemeyer D.L., Menke J.G., Martin-Gallardo A., Hughes J.V.,  
RA Young A., Mitra S.W.;  
RT "Molecular cloning and partial sequencing of hepatitis A viral cDNA."  
RL J. Virol. 54:247-255(1985).  
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M10033; AAA45470.1; -.  
DR PIR; A03904; GNNYHA.  
DR InterPro; IPR008975; Viral_cap_coat.  
KW Polyprotein; Coat protein; Core protein.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 245 981 CORE PROTEIN P2A.  
FT CHAIN 981 1076 CORE PROTEIN P2B.
```

```
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 836 COAT PROTEIN VP1 (PID).  
FT CHAIN 837 >852 CORE PROTEIN P2A.  
FT NON_TER 852 852  
SQ SEQUENCE 852 AA; 95563 MW; 73D3ED0AD532820E CRC64;  
  
Query Match 97.2%; Score 106; DB 1; Length 852;  
Best Local Similarity 95.0%; Pred. No. 1.7e-08;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 779 VDDPRSEEDKRFESHIECRK 798  
[1]  
-----  
RESULT 8  
POLG_HPAVL STANDARD; PRT; 2227 AA.  
AC P06431;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain LA).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI_TaxID=12099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85190549; PubMed=2986127;  
RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,  
RA Merryweather J., van Nest G., Dina D.;  
RT "Primary structure and gene organization of human hepatitis A virus."  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC {RNA} (N).  
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; K02990; AAA45472.1; -.  
DR PIR; A03903; GNNYHR.  
DR MEROPS; C03.005; -.  
DR InterPro; IPR004004; Calici_pol_hel.  
DR InterPro; IPR009003; Cys_Ser_trypsin.  
DR InterPro; IPR000605; RNA_helicase.  
DR InterPro; IPR007095; RNA_pol_DS_PS.  
DR InterPro; IPR001205; RNA_pol_P3D.  
DR InterPro; IPR007094; RNA_pol_PSVir.  
DR InterPro; IPR008975; Viral_cap_coat.  
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.  
DR Pfam; PF00910; RNA_helicase; 1.  
DR PRINTS; PK00918; CALICIVIRUSN.  
DR Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 836 COAT PROTEIN VP1 (PID).  
FT CHAIN 837 980 CORE PROTEIN P2A.  
FT CHAIN 981 1076 CORE PROTEIN P2B.
```

FT CHAIN 1077 1422 CORE PROTEIN P2C.  
 FT CHAIN 1423 1484 PROBABLE PROTEIN P3A.  
 FT CHAIN 1485 1507 PROBABLE PROTEIN P3B.  
 FT CHAIN 1508 1678 PROBABLE PROTEIN P3C.  
 FT CHAIN 1679 2227 RNA-DIRECTED POLYMERASE P3D.  
 SQ SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;

Query Match 97.2%; Score 106; DB 1; Length 2227;  
 Best Local Similarity 95.0%; Pred. No. 4.9e-08;  
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEDKRFESHIECRK 20  
 |||||:|||||  
 Db 779 VDDPRSEDKRFESHIECRK 798

RESULT 9  
 POLG\_HPAVG STANDARD; PRT; 808 AA.  
 AC Q02381;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein P2A] (Fragment).  
 DE Hepatitis A virus (strain GA76).  
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 CC NCBI\_TaxID=31706;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92260183; PubMed=13116423;  
 RA Khanna B., Spelbring J.E., Innis B.L., Robertson B.H.;  
 RT "Characterization of a genetic variant of human hepatitis A virus.";  
 RL J. Med. Virol. 36:118-124 (1992).  
 CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 each of which is composed of one copy each of proteins VP1, VP2,  
 VP3, and VP4.  
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M66595; AAA45477.1; -;  
 DR InterPro; IPR008975; Viral\_cap.coat.  
 KW Polyprotein; Coat protein; Core protein.  
 FT NON\_TER 1 1  
 FT CHAIN <1 2 COAT PROTEIN VP4 (PIA).  
 FT CHAIN 3 223 COAT PROTEIN VP2 (PIB).  
 FT CHAIN 224 470 COAT PROTEIN VP3 (PIC).  
 FT CHAIN 471 770 COAT PROTEIN VP1 (PID).  
 FT CHAIN 771 >808 CORE PROTEIN P2A.  
 FT NON\_TER 808 808  
 SQ SEQUENCE 808 AA; 90632 MW; D80CE7E57A479C12 CRC64;

Query Match 80.7%; Score 88; DB 1; Length 808;  
 Best Local Similarity 80.0%; Pred. No. 9.4e-06;  
 Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDDPRSEDKRFESHIECRK 20  
 |||||:|||||  
 Db 758 VDDPRTEEDKRFESHIECRK 777

RESULT 10  
 POLG\_HPAVS STANDARD; PRT; 2230 AA.  
 ID POLG\_HPAVS  
 AC P14553;

DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
 DE P3D (EC 2.7.7.48)].  
 OS Simian hepatitis A virus (strain AGM-27).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 CC NCBI\_TaxID=12102;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91311420; PubMed=1649901;  
 RA Tsarev S.A., Emerson S.U., Balayan M.S., Ticehurst J.R.,  
 RA Purcell R.H.;  
 RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome  
 structure and growth in cell culture with other HAV strains.";  
 RL J. Gen. Virol. 72:1677-1683 (1991).  
 RN [2]  
 RP SEQUENCE OF 1750-2164 FROM N.A.  
 RX MEDLINE=89232168; PubMed=2541023;  
 RA Balayan M.S., Kusov Y.Y., Andjaparidze A.G., Tsarev S.A.,  
 RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;  
 RT "Variations in genome fragments coding for RNA polymerase in human  
 and simian hepatitis A viruses.";  
 RL FEBS Lett. 247:425-428 (1989).  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 {RNA}(N).  
 CC -!- SUBUNIT: The virus capsid is composed of one copy each of proteins VP1, VP2,  
 each of which is composed of one copy each of proteins VP1, VP2,  
 VP3, and VP4.  
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; D00924; BAA00766.1; -;  
 DR EMBL; X15461; CAA33490.1; -;  
 DR PIR; A30470; GNNYSA.  
 DR MEROPS; C03.005; -;  
 DR InterPro; IPR004004; Calici\_pol\_hel.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR006005; RNA\_helicase.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR001205; RNA\_pol\_P3D.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR InterPro; IPR008975; Viral\_cap.coat.  
 DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR Pfam; PF00910; RNA\_helicase; 1.  
 DR PRINTS; PR00918; CALICIVIRUS.  
 KW Polyprotein; Coat protein; Core protein; Transferase;  
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
 FT CHAIN 1 27 COAT PROTEIN VP4 (PIA).  
 FT CHAIN 28 249 COAT PROTEIN VP2 (PIB).  
 FT CHAIN 250 495 COAT PROTEIN VP3 (PIC).  
 FT CHAIN 496 795 COAT PROTEIN VP1 (PID).  
 FT CHAIN 796 984 CORE PROTEIN P2A.  
 FT CHAIN 985 1091 CORE PROTEIN P2B.  
 FT CHAIN 1092 1426 CORE PROTEIN P2C.  
 FT CHAIN 1427 1498 PROBABLE PROTEIN 3A.  
 FT CHAIN 1499 1521 PROBABLE PROTEIN 3B.  
 FT CHAIN 1522 1741 PROBABLE PROTEIN 3C.  
 FT CHAIN 1742 2230 RNA-DIRECTED POLYMERASE 3D.  
 SQ SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;

Query Match 77.1%; Score 84; DB 1; Length 2230;  
 Best Local Similarity 94.1%; Pred. No. 0.00012;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIE 17  
|||||  
DB 783 VDDPRSEEDKRFESHIE 799

## RESULT 11

POLG\_HPAVT  
ID POLG\_HPAVT STANDARD; PRT; 839 AA.  
AC P31788;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein P2A] (Fragment).  
DE P2A)  
OS Simian hepatitis A virus (strain CV-145).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=31707;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91311421; PubMed=1649902;  
RA Nainan O.V., Margolis H.S., Robertson B.H., Balayan M., Brinton M.A.;  
RT "Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus macaques (Macaca fascicularis).";  
RL J. Gen. Virol. 72:1685-1689 (1991).  
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units, each of which is composed of one copy each of proteins VP1, VP2, VP3, and VP4.  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M59286; AAA45473.1; -;  
DR PIR; JQ1180; GNNYS2.  
DR InterPro; IPR008975; Viral cap coat.  
KW Polyprotein; Coat protein; Core protein.  
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
FT CHAIN 492 ? COAT PROTEIN VP1 (P1D).  
FT CHAIN ? >839 CORE PROTEIN P2A.  
FT NON\_TER 839 839  
SQ SEQUENCE 839 AA; 93825 MW; 2CACCA4BD1E192DBC CRC64;

Query Match 74.3%; Score 81; DB 1; Length 839;  
Best Local Similarity 82.4%; Pred. No. 0.00012;  
Matches 14; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIE 17  
|||||  
DB 779 VDDPRSEEDKRFESHIE 795

## RESULT 12

CO3\_RAT  
ID CO3\_RAT STANDARD; PRT; 1663 AA.  
AC P01026;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Complement C3 precursor [Contains: C3a anaphylatoxin].  
DE C3.  
GN C3.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

NCBI\_TaxID=101116;  
[1]  
SEQUENCE FROM N.A.  
RX STRAIN=Wistar; TISSUE=Liver;  
RX MEDLINE=90245672; PubMed=2336397;  
RA Misumi Y., Sohma M., Ikehara Y.;  
RT "Nucleotide and deduced amino acid sequence of rat complement C3.";  
RL Nucleic Acids Res. 18:2178-2178 (1990).  
[2]  
SEQUENCE OF 671-748.  
RX MEDLINE=79062262; PubMed=309768;  
RA Jacobs J.W., Rubin J.S., Hugli T.E., Bogardt R.A., Mariz I.K.,  
RA Daniels J.S., Daughaday W.H., Bradshaw R.A.;  
RT "Purification, characterization, and amino acid sequence of rat anaphylatoxin (C3a).";  
RL Biochemistry 17:5031-5038 (1978).  
[3]  
SEQUENCE OF 1316-1595 FROM N.A.  
RX MEDLINE=89380332; PubMed=2674144;  
RA Sundstrom S.A., Komm B.S., Ponce-De-Leon H., Yi Z., Teuscher C.,  
RA Lyttle C.R.;  
RT "Estrogen regulation of tissue-specific expression of complement C3";  
J. Biol. Chem. 264:16941-16947 (1989).  
CC -!- FUNCTION: C3 plays a central role in the activation of the complement system. Its processing by C3 convertase is the central reaction in both classical and alternative complement pathways. After activation C3b can bind covalently, via its reactive thioester, to cell surface carbohydrates or immune aggregates.  
CC -!- FUNCTION: Derived from proteolytic degradation of complement C3, C3a anaphylatoxin is a mediator of local inflammatory process. It induces the contraction of smooth muscle, increases vascular permeability and causes histamine release from mast cells and basophilic leukocytes.  
CC -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN, RELEASING C3a ANAPHYLATOXIN & GENERATING C3b (BETA CHAIN + ALPHA CHAIN).  
CC -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.  
CC -!- SIMILARITY: Contains 1 anaphylatoxin-like domain.  
CC -!- SIMILARITY: Contains 1 NTR domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X52477; CAA36716.1; -;  
DR EMBL; M28866; AAA40837.1; ALT\_SEQ.  
DR PIR; S15764; C3RT.  
DR PDB; 1QQF; 31-JUL-00.  
DR PDB; 1QSU; 31-JUL-00.  
DR InterPro; IPR002890; A2M\_N.  
DR InterPro; IPR009048; AM receptor bind.  
DR InterPro; IPR000020; Anaphylatoxin.  
DR InterPro; IPR001840; Anaphylatoxin.  
DR InterPro; IPR001599; MacrogloblnA2.  
DR InterPro; IPR001134; Netrin\_C.  
DR InterPro; IPR008930; Terp\_cyc\_toroid.  
DR Pfam; PF00207; A2M; 1.  
DR Pfam; PF01835; A2M\_N; 1.  
DR Pfam; PF01821; ANATO; 1.  
DR Pfam; PF01759; NTR; 1.  
DR PRINTS; PR000004; ANAPHYLATOXN.  
DR ProDom; PD003264; Anaphylatoxin; 1.  
DR SMART; SM00104; ANATO; 1.  
DR SMART; SM00643; C345C; 1.  
DR PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; 1.  
DR PROSITE; PS01177; ANAPHYLATOXIN\_1; 1.  
DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.

complement system. Its processing by C3 convertase is the central reaction in both classical and alternative complement pathways. After activation C3b can bind covalently, via its reactive thioester, to cell surface carbohydrates or immune aggregates. C3a anaphylatoxin is a mediator of local inflammatory process. It induces the contraction of smooth muscle, increases vascular permeability and causes histamine release from mast cells and basophilic leukocytes.

-!- SUBUNIT: C3 precursor is first processed by the removal of 4 Arg residues, forming two chains, beta and alpha, linked by a disulfide bond. C3 convertase activates C3 by cleaving the alpha chain, releasing C3a anaphylatoxin and generating C3b (beta chain + alpha' chain).

-!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.

-!- SIMILARITY: Contains 1 anaphylatoxin-like domain.

-!- SIMILARITY: Contains 1 NTR domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL; M34054; AAA37038.1; -  
PIR; A37156; C3GP.  
HSSP; P01024; 1C3D.  
InterPro; IPR002890; A2M N.  
InterPro; IPR009048; AM\_receptor\_bind.  
InterPro; IPR000020; Anaphylatoxin.  
InterPro; IPR001840; Anaphylatoxn.  
InterPro; IPR001599; MacrogloblnA2.  
InterPro; IPR001134; Netrin C.  
InterPro; IPR008930; Terp\_cyc\_toroid.  
Pfam; PF00207; A2M; 1.  
Pfam; PF01835; A2M N; 1.  
Pfam; PF01821; ANATO; 1.  
Pfam; PF01759; NTR; 1.  
PRINTS; PR00004; ANAPHYLATOXN.  
ProDom; PD003264; Anaphylatoxin; 1.  
SMART; SM00104; ANATO; 1.  
SMART; SM00643; C345C; 1.  
PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; 1.  
PROSITE; PS01177; ANAPHYLATOXIN\_1; 1.  
PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.  
PROSITE; PS0189; NTR; 1.  
Complement pathway; Complement alternate pathway; Plasma; Inflammatory response; Glycoprotein; Signal; Thioester bond.  
KW SIGNAL 1 29  
FT CHAIN 30 1666  
FT CHAIN 30 671  
FT CHAIN 676 1666  
FT PEPTIDE 676 753  
FT CHAIN 754 1666  
FT DOMAIN 698 733  
FT DOMAIN 1522 1664  
FT SITE 753 754  
FT SITE 821 754  
FT DISULFID 557 821  
FT DISULFID 630 666  
FT DISULFID 698 725  
FT DISULFID 698 732  
FT DISULFID 712 733  
FT DISULFID 878 1517  
FT DISULFID 1106 1163  
FT DISULFID 1363 1493  
FT DISULFID 1394 1462  
FT DISULFID 1510 1515  
FT DISULFID 1522 1593  
FT DISULFID 1522 1664  
FT DISULFID 1640 1649  
FT DISULFID 1640 1649  
FT CROSSLINK 1015 1018  
Isoglutamyl cysteine thioester (Cys-Gln).

PROSITE; PS0189; NTR; 1.  
Complement pathway; Complement alternate pathway; Plasma; Inflammatory response; Glycoprotein; Signal; 3D-structure; Thioester bond.  
FT SIGNAL 1 24  
FT CHAIN 25 1663  
FT CHAIN 25 666  
FT CHAIN 671 1663  
FT CHAIN 671 1663  
FT PEPTIDE 671 748  
FT CHAIN 749 1663  
FT DOMAIN 693 728  
FT DOMAIN 1518 1661  
FT SITE 748 749  
FT SITE 816 749  
FT DISULFID 558 816  
FT DISULFID 626 661  
FT DISULFID 693 720  
FT DISULFID 693 720  
FT DISULFID 694 727  
FT DISULFID 707 728  
FT DISULFID 873 1513  
FT DISULFID 1101 1158  
FT DISULFID 1358 1489  
FT DISULFID 1389 1458  
FT DISULFID 1506 1511  
FT DISULFID 1518 1590  
FT DISULFID 1537 1661  
FT DISULFID 1537 1661  
FT CROSSLINK 1010 1013  
FT CARBOHYD 939 939  
FT CARBOHYD 1617 1617  
FT CONFLICT 721 722  
FT CONFLICT 721 722  
SEQUENCE 1663 AA; 2F87CCB143CD4BC CRC64;  
Query Match 45.9%; Score 50; DB 1; Length 1663;  
Best Local Similarity 38.9%; Pred. No. 14;  
Matches 7; Conservative 8; Mismatches 3; Indels 0; Gaps 0;  
LK -> KL (IN REF. 2).

QY 2 DDPRSEEDKRFESHEICR 19  
D5 1574 DEVQAGQRRRISHVKCR 1591

RESULT 13  
CO3\_CAVPO  
ID - CO3\_CAVPO STANDARD; PRT; 1666 AA.  
AC P12387;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Complement C3 precursor [Contains: C3a anaphylatoxin].  
GN C3.  
OS Cavia porcellus (Guinea pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
OX NCBI\_TaxID=10141;  
RN [1]\_TaxID=10141;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90307998; PubMed=1973176;  
RA Auerbach H.S., Burger R., Dodds A., Colten H.R.;  
RT "Molecular basis of complement C3 deficiency in guinea pigs.";  
RL J. Clin. Invest. 86:96-106(1990).  
RN [2]  
RP SEQUENCE OF 676-753.  
RX MEDLINE=89113342; PubMed=3064079;  
RA Gerard N.P., Lively M.O., Gerard C.;  
RT "Amino acid sequence of guinea pig C3a anaphylatoxin.";  
RL Protein Seq. Data Anal. 1:473-478(1988).  
RN [3]  
RP SEQUENCE OF 993-1032.  
RX MEDLINE=83178889; PubMed=6838833;  
RA Thomas M.L., Tack B.F.;  
RT "Identification and alignment of a thiol ester site in the third component of guinea pig complement.";  
RL Biochemistry 22:942-947(1983).  
RL Biochemistry 22:942-947(1983).  
-!- FUNCTION: C3 plays a central role in the activation of the





Db 42 DDRAVEENRRRLTERLECR 61

RESULT 16

CC3 MOUSE STANDARD; PRT; 1663 AA.

AD CO3 MOUSE P01027;

AC 21-JUL-1996 (Rel. 01. Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Complement C3 precursor (HSE-MSF) [Contains: C3a anaphylatoxin].

GN C3.

OS Mus musculus (Mouse).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM LONG).

RX MEDLINE=8503854; PubMed=6208565;

RA Fey G.H., Lundwall A., Wetsel R.A., Tack B.F., de Bruijn M.H.L.,

RA Domdey H.;

RT "Nucleotide sequence of complementary DNA and derived amino acid

RT sequence of murine complement protein C3.";

RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 306:333-344(1984).

RN [2]

RP SEQUENCE OF 671-1663 FROM N.A. (ISOFORM LONG).

RX MEDLINE=85054819; PubMed=6094532;

RA Wetsel R.A., Lundwall A., Davidson F., Gibson T., Tack B.F., Fey G.H.;

RT "Structure of murine complement component C3. II. Nucleotide sequence

RT of cloned complementary DNA coding for the alpha chain.";

RL J. Biol. Chem. 259:13857-13862(1984).

RN [3]

RP SEQUENCE OF 671-748 FROM N.A.

RX MEDLINE=83117730; PubMed=6961437;

RA Domdey H., Wiebauer K., Kazmaier M., Mueller V., Odink K., Fey G.H.;

RT "Characterization of the mRNA and cloned cDNA specifying the third

RT component of mouse complement.";

RL Proc. Natl. Acad. Sci. U.S.A. 79:7619-7623(1982).

RN [4]

RP SEQUENCE OF 658-761 FROM N.A.

RX MEDLINE=84201365; PubMed=6609661;

RA Fey G.H., Wiebauer K., Domdey H.;

RT "Amino acid sequences of mouse complement C3 derived from nucleotide

RT sequences of cloned cDNA.";

RL Ann. N.Y. Acad. Sci. 421:307-312(1983).

RN [5]

RP SEQUENCE OF 1-34 FROM N.A.

RX MEDLINE=83117622; PubMed=6985486;

RA Wiebauer K., Domdey H., Digelmann H., Fey G.;

RT "Isolation and analysis of genomic DNA clones encoding the third

RT component of mouse complement.";

RL Proc. Natl. Acad. Sci. U.S.A. 79:7077-7081(1982).

RN [6]

RP SEQUENCE OF 25-41 AND 749-760.

RX MEDLINE=93373334; PubMed=8364938;

RA Hamada J.-I., Cavanaugh P.G., Miki K., Nicolson G.L.;

RT "A paracrine migration-stimulating factor for metastatic tumor cells

RT secreted by mouse hepatic sinusoidal endothelial cells:

RT identification as complement component C3b.";

RL Cancer Res. 53:4418-4423(1993).

RN [7]

RP ALTERNATIVE INITIATION.

RX MEDLINE=95053742; PubMed=7964485;

RA Cahen-Kramer Y., Martenson I.L., Melchers P.;

RT "The structure of an alternate form of complement C3 that displays

RT costimulatory growth factor activity for B lymphocytes.";

RL J. Exp. Med. 180:2079-2088(1994).

CC -1- FUNCTION: C3 plays a central role in the activation of the

CC complement system. Its processing by C3 convertase is the central

CC reaction in both classical and alternative complement pathways.

CC After activation C3b can bind covalently, via its reactive

CC thioester, to cell surface carbohydrates or immune aggregates.

CC -1- FUNCTION: Derived from proteolytic degradation of complement C3,

CC

CC C3a anaphylatoxin is a mediator of local inflammatory process. It

CC induces the contraction of smooth muscle, increases vascular

CC permeability and causes histamine release from mast cells and

CC basophilic leukocytes. The short isoform has B-cell stimulatory

CC activity.

CC -1- SUBUNIT: C3 precursor is first processed by the removal of 4 Arg

CC residues, forming two chains, beta and alpha, linked by a

CC disulfide bond. C3 convertase activates C3 by cleaving the alpha

CC chain, releasing C3a anaphylatoxin and generating C3b (beta chain

CC + alpha' chain)

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative initiation;

CC Comment=2 isoforms, Long (shown here) and Short, are produced by

CC alternative initiation;

CC -1- PTM: C3b is rapidly split in two positions by factor I and a

CC cofactor to form iC3b (inactivated C3b) and C3f which is released.

CC Then iC3b is slowly cleaved (possibly by factor I) to form C3c

CC and C3dg. Other proteases produce other fragments such as C3d or

CC C3g.

CC -1- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.

CC -1- SIMILARITY: Contains 1 anaphylatoxin-like domain.

CC -1- SIMILARITY: Contains 1 NTR domain.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC

DR EMBL; K02782; AAC42013.1; -

DR EMBL; J00369; AAA37336.1; -

DR EMBL; J00367; AAA37336.1; JOINED.

DR EMBL; M33032; AAA37378.1; -

DR EMBL; Z37998; CAA86099.2; -

DR PIR; A92459; C3MS.

DR HSSP; P01024; 1C3D.

DR MGD; MGI:88227; C3.

DR InterPro; IPR002890; A2M\_N.

DR InterPro; IPR009048; AM\_receptor\_bind.

DR InterPro; IPR000020; Anaphylatoxin.

DR InterPro; IPR001840; Anaphylatoxn.

DR InterPro; IPR001599; MacroglobinA2.

DR InterPro; IPR001134; Netrin\_C.

DR InterPro; IPR008930; Terp\_cyc\_toroid.

DR Pfam; PF0207; A2M; 1.

DR Pfam; PF01835; A2M\_N; 1.

DR Pfam; PF01821; ANATO; 1.

DR Pfam; PF01759; NTR; 1.

DR PRINTS; PR00004; ANAPHYLATOXN.

DR ProDom; PD003264; Anaphylatoxin; 1.

DR SMART; SM00104; ANATO; 1.

DR SMART; SM00643; C345C; 1.

DR PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; 1.

DR PROSITE; PS01177; ANAPHYLATOXIN\_1; 1.

DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.

DR PROSITE; PS0189; NTR; 1.

KW Complement pathway; Complement alternate pathway; Plasma;

KW Inflammatory response; Glycoprotein; Signal; Alternative initiation;

KW Thioester bond.

FT SIGNAL 1 24

FT CHAIN 25 1663 COMPLEMENT C3, ISOFORM LONG.

FT CHAIN 25 666 COMPLEMENT C3, BETA CHAIN.

FT CHAIN 671 1663 COMPLEMENT C3, ALPHA CHAIN.

FT CHAIN 1129 1663 COMPLEMENT C3, ISOFORM SHORT.

FT INIT\_MET 1129 1129 FOR ISOFORM SHORT.

FT PEPTIDE 671 748 C3A ANAPHYLATOXIN.

FT CHAIN 749 1663 C3B (ALPHA' CHAIN).

FT PEPTIDE 749 954 C3C FRAGMENT.

FT PEPTIDE 955 1303 C3DG FRAGMENT.

FT PEPTIDE 955 1001 C3G FRAGMENT.

FT PEPTIDE 1002 1303 C3D FRAGMENT.

PEPTIDE 1304 1320 C3F FRAGMENT.  
FT DOMAIN 693 728 ANAPHYLATOXIN-LIKE.  
FT DOMAIN 1518 1661 NTR.  
FT SITE 748 749 CLEAVAGE (BY C3 CONVERTASE).  
FT SITE 1303 1304 CLEAVAGE (BY FACTOR I).  
FT SITE 1320 1321 CLEAVAGE (BY FACTOR I).  
FT SITE 559 616 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 626 661 BY SIMILARITY.  
FT DISULFID 693 720 BY SIMILARITY.  
FT DISULFID 694 727 BY SIMILARITY.  
FT DISULFID 707 728 BY SIMILARITY.  
FT DISULFID 873 1513 BY SIMILARITY.  
FT DISULFID 1101 1158 BY SIMILARITY.  
FT DISULFID 1358 1489 BY SIMILARITY.  
FT DISULFID 1389 1458 BY SIMILARITY.  
FT DISULFID 1506 1511 BY SIMILARITY.  
FT DISULFID 1518 1590 BY SIMILARITY.  
FT DISULFID 1537 1661 BY SIMILARITY.  
FT DISULFID 1637 1646 BY SIMILARITY.  
FT CARBOHYD 939 939 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 1617 1617 N-LINKED (GLCNAC. . .).  
FT CROSSLINK 1010 1013 Isoglutamyl cysteine thioester (Cys-Gln)  
SQ SEQUENCE 1663 AA; 186482 MW; DE5546CC769BEA19 CRC64;  
Query Match 42.2%; Score 46; DB 1; Length 1663;  
Best Local Similarity 38.9%; Pred. No. 58;  
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;  
QY 2 DDPRSEEDKRFESHIECR 19  
DB 1574 DEVQAGQKRFISHIKCR 1591  
RESULT 17  
YD35 MYCPN  
ID YD35 MYCPN STANDARD; PRT; 741 AA.  
AC P75443;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein MPN335 (F10\_orf741).  
GN MPN335 OR WP501.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129;  
RX MEDLINE=97105885; PubMed=8948633;  
RA Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,  
RA Herrmann R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae";  
RL Nucleic Acids Res. 24:4420-4449(1996).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -1- SIMILARITY: TO M.PNEUMONIAE MPN333.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; A300049; AAB96149.1; -  
DR DR PIR; S73827; S73827.  
DR KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRANSMEM 34 54 POTENTIAL.  
FT TRANSMEM 76 96 POTENTIAL.  
FT TRANSMEM 120 140 POTENTIAL.  
FT TRANSMEM 156 176 POTENTIAL.

FT TRANSMEM 187 207 POTENTIAL.  
FT TRANSMEM 655 675 POTENTIAL.  
FT TRANSMEM 715 735 POTENTIAL.  
SQ SEQUENCE 741 AA; 85025 MW; 1776A96BCF83567F CRC64;  
Query Match 41.3%; Score 45; DB 1; Length 741;  
Best Local Similarity 42.1%; Pred. No. 34;  
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
QY 2 DDPRSEEDKRFESHIECR 20  
DB 629 EDKQSEEEKFDEIENAK 647  
RESULT 18  
RL4 CHLTR  
ID RL4 CHLTR STANDARD; PRT; 222 AA.  
AC Q84532;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 50S ribosomal protein L4.  
GN RPLD OR RL4 OR CTS27.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D/UW-3/Cx;  
RX MEDLINE=9900809; PubMed=97841136;  
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;  
RA "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis";  
RL Science 282:754-759(1998).  
CC -1- FUNCTION: This protein binds directly and specifically to 23S rRNA (By similarity).  
CC -1- SIMILARITY: Belongs to the L4P family of ribosomal proteins.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; AE001323; AAC68128.1; -  
DR DR PIR; E71507; E71507.  
DR InterPro; IPR002136; Ribosomal\_L4/L1E.  
DR Pfam; PF00573; Ribosomal\_L4; 1.  
DR Ribosomal protein; rRNA-binding; Complete proteome.  
KW Ribosomal protein; rRNA-binding; Complete proteome.  
SQ SEQUENCE 222 AA; 24506 MW; 44641751527630D1 CRC64;  
Query Match 40.4%; Score 44; DB 1; Length 222;  
Best Local Similarity 42.9%; Pred. No. 13;  
Matches 9; Conservative 6; Mismatches 4; Indels 2; Gaps 1;  
QY 1 VDDPRSEEDKRF--ESHIECR 19  
DB 135 LDAPKTKKALRFLKCNVECR 155  
RESULT 19  
TRMB\_HUMAN  
ID TRMB\_HUMAN STANDARD; PRT; 276 AA.  
AC Q9UBP6; Q14105;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE tRNA (guanine-N(7)-)-methyltransferase (EC 2.1.1.33) (tRNA(m7G46)-methyltransferase) (Methyltransferase-like protein 1).

GN METTL1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
RP TISSUE=Brain;  
RX MEDLINE=99263505; PubMed=10329009;  
RA Bahr A., Hankeln T., Fiedler T., Hegemann J., Schmidt E.R.;  
RT "Molecular analysis of METTL1, a novel human methyltransferase-like  
RT gene with a high degree of phylogenetic conservation.";  
RL Genomics 57:424-428(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan K.F., Moore T., Max S.I., Wang J., Heist F.,  
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP FUNCTION, AND INTERACTION WITH WDR4.  
RX MEDLINE=22290633; PubMed=12403464;  
RA Alexandrov A., Martzen M.R., Phizicky E.M.;  
RT "Two proteins that form a complex are required for 7-methylguanosine  
RT modification of yeast tRNA.";  
RL RNA 8:1253-1266(2002).  
CC -!- FUNCTION: Catalyzes the formation of N(7)-methylguanine at  
CC position 46 (m7G46) in tRNA.  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-  
CC homocysteine + tRNA containing N(7)-methylguanine.  
CC -!- SUBUNIT: Forms a complex with WDR4.  
CC -!- TISSUE SPECIFICITY: Ubiquitous.  
CC -!- SIMILARITY: Belongs to the methyltransferase superfamily. Trmb  
CC family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; Y18643; CAAT7239.1; --  
CC EMBL; Y18642; CAAT7238.1; --  
CC EMBL; X96698; CAA65470.1; ALT\_INIT.  
CC EMBL; BC000550; AAH00550.1; --  
CC Genew; HGNC:7030; METTL1.  
CC MIM; 604465; --  
CC GO; GO:0008169; F:methyltransferase activity; TAS.  
CC GO; GO:0008176; F:tRNA (guanine-N7-)-methyltransferase activity; IDA.  
CC GO; GO:0006400; P:tRNA modification; IDA.  
CC InterPro; IPR004395; Cons hypoth91.  
CC InterPro; IPR003358; Methyltransf\_4.  
CC InterPro; IPR000051; SAM\_bind.

DR Pfam; PF02390; Methyltransf 4; 1.  
KW TIGRFAW; TIGR00091; TIGR00091; 1.  
KW Transferase; Methyltransferase; tRNA processing.  
SQ SEQUENCE 276 AA; 31471 MW; 715AE65A18632892 CRC64;  
Query Match 40.4%; Score 44; DB 1; Length 276;  
Best Local Similarity 37.5%; Pred. No. 17;  
Matches 6; Conservative 7; Mismatches 3; Indels 0; Gaps 0;  
QY 2 DDPSEEDKRFESHIE 17  
DB 64 DDPKDKKRAQAVE 79  
RESULT 20  
FURC\_DROME STANDARD; PRT; 1101 AA.  
ID FURC\_DROME STANDARD; PRT; 1101 AA.  
AC P30430; Q27235; Q9VBR5;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Furin-like protease 1, isoform 1-CRR precursor (EC 3.4.21.75) (Furin  
DE 1) (Kex2-like endoprotease 1) (dKLP-1).  
GN FUR1 OR CGI0772.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Oregon-R; TISSUE=Head;  
RX MEDLINE=92185516; PubMed=1545235;  
RA Hayflick J.S., Wolfgang W.J., Forte M.A., Thomas G.;  
RT "A unique Kex2-like endoprotease from Drosophila melanogaster is  
RT expressed in the central nervous system during early embryogenesis.";  
RL J. Neurosci. 12:705-717(1992).  
RN [2]  
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND  
RP DEVELOPMENTAL STAGE.  
RC STRAIN=Tuebingen, and Oregon-R; TISSUE=Embryo;  
RX MEDLINE=93259127; PubMed=8491178;  
RA Roebroek A.J.M., Creemers J.W.M., Pauli I.G.L., Bogaert T.,  
RA Van de Ven W.J.M.;  
RT "Generation of structural and functional diversity in furin-like  
RT proteins in Drosophila melanogaster by alternative splicing of the  
RT Fur1 gene.";  
RL EMBL J. 12:1853-1870(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.B.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadeu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Flesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Weissman D.A., Weinstock G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195 (2000).  
RN [4]  
RP REVISIONS, AND ALTERNATIVE SPLICING.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S., Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E., Smith C.D., Tupy J.L., Whifield E.J., Bayraktaroglu L., Berman B.P., Bettencourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A., Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q., Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M., Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
CC -!- FUNCTION: Furin is likely to represent the ubiquitous endoprotease activity within constitutive secretory pathways and capable of cleavage at the RX(K/R)R consensus motif (by similarity).  
CC -!- CATALYTIC ACTIVITY: Release of mature proteins from their proproteins by cleavage of Arg-Xaa-Yaa-Arg-|-Zaa bonds, where Xaa can be any amino acid and Yaa is Arg or Lys. Releases albumin, complement component C3 and von Willebrand factor from their respective precursors.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Golgi.  
CC -!- ALTERNATIVE PRODUCTS.  
CC Event-Alternative splicing; Named isoforms=4;  
CC Comment-Experimental confirmation may be lacking for some isoforms;  
CC Name=1-CRR; Synonyms=A;  
CC IsoId=P30430-1; Sequence=Displayed;  
CC Name=1; Synonyms=F;  
CC IsoId=P26016-2; Sequence=External;  
CC Name=1-X; Synonyms=B;  
CC IsoId=P26016-1; Sequence=External;  
CC Name=2; Synonyms=C, D, E;  
CC IsoId=P26016-3; Sequence=External;  
CC -!- TISSUE SPECIFICITY: In adults, isoform 1-CRR is expressed in CNS, fat body, and female reproductive tissues, and in embryos, in anal pads, hindgut, developing antennomaxillary complex, oenocytes, clypeolabrum, pharynx, trachea, CNS and developing posterior spiracles.  
CC -!- DEVELOPMENTAL STAGE: Isoform 1-CRR is expressed in embryos, larvae, pupae and adults.  
CC -!- SIMILARITY: Belongs to peptidase family S8. Furin subfamily.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M81431; AAA28467.1; -;  
CC EMBL; L12370; AAA28545.1; -;  
CC EMBL; L12369; AAA28545.1; JOINED.  
CC EMBL; L12372; AAA28548.1; -;  
CC EMBL; AB003751; AANI4051.1; -;

DR MEROPS; S08.048; -;  
DR FlyBase; FBgn0004509; Furl.  
DR GO; GO:0005886; C:plasma membrane; IDA.  
DR GO; GO:0004276; F:furin activity; IDA.  
DR InterPro; IPR006212; Furin\_repeat.  
DR InterPro; IPR002029; Peptidase\_S8.  
DR InterPro; IPR002884; Peptidase\_S8B.  
DR InterPro; IPR009020; Protease\_inhib.  
DR Pfam; PF01483; P\_protease; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR PRINTS; PD00723; SUBTILISIN.  
DR PRODOM; PD000717; P\_domain; 1.  
DR SMART; SMC0261; FU; 2.  
DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
KW Hydrolyase; Serine protease; Glycoprotein; Signal; Transmembrane;  
KW Multigene family; Zymogen; Alternative splicing; Golgi stack.  
FT SIGNAL 1; POTENTIAL.  
FT PROPEP ? 309 POTENTIAL.  
FT CHAIN 310 1101 FURIN-LIKE PROTEASE 1, ISOFORM 1-CRR.  
FT DOMAIN 310 655 CATALYTIC.  
FT ACT\_SITE 372 372 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 413 413 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 587 587 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT TRANSMEM 119 139 POTENTIAL.  
FT DOMAIN 1014 1034 POTENTIAL.  
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 426 426 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 606 606 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 727 727 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 859 859 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 1014 1014 T -> I (IN REF. 1).  
SQ SEQUENCE 1101 AA; 120993 MW; 90DC38E2CACB71A0 CRC64;  
Query Match 40.4%; Score 44; DB 1; Length 1101;  
Best Local Similarity 42.1%; Pred. No. 75;  
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
QY 2 DDPSEEDKRFESHIECRK 20  
Db 1057 DESDSEDELFTHFPARK 1075  
RESULT 21  
DHGA ACICA STANDARD; PRT; 801 AA.  
ID DHGA ACICA STANDARD; PRT; 801 AA.  
AC P05465;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Glucose dehydrogenase-A [pyrroloquinoline-quinone] precursor  
DE (EC 1.1.99.17) (Quinoprotein glucose DH) (GDH-A).  
GN GDHA.  
OS Acinetobacter calcoaceticus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Moraxellaceae; Acinetobacter.  
OX NCBI\_TaxID=471;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LMD 79.41;  
RX MEDLINE=88289368; PubMed=3399393;  
RA Clifton-Jensen A.-M., Goosen N., Odle G., van de Putte P.;  
RT "Nucleotide sequence of the gene coding for quinoprotein glucose dehydrogenase from Acinetobacter calcoaceticus.";  
RL Nucleic Acids Res. 16:6228-6228(1988).  
CC -!- CATALYTIC ACTIVITY: D-glucose + acceptor = D-glucono-1,5-lactone +

CC reduced acceptor.  
 CC -1- COFACTOR: PQQ.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- MISCELLANEOUS: ACINETOBACTER CALCOACETICUS CONTAINS TWO DIFFERENT  
 CC PQQ DEPENDENT ENZYMES WITH GDH ACTIVITY. GDH-A PREFERENCES  
 CC 2-DEOXYGLUCOSE AS SUBSTRATE, THE SPECIFIC SUBSTRATES FOR GDH-B ARE  
 CC DISACCHARIDES.  
 CC -1- SIMILARITY: Belongs to the bacterial PQQ dehydrogenase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X07235; CAA30222.1; -.  
 CC PIR; S00943; S00943.  
 CC InterPro; IPR001479; Bac\_PQQ.  
 CC InterPro; IPR002372; Bac\_PQQ\_repeat.  
 CC Pfam; PF01011; PQQ; 4.  
 CC SMART; SM00564; PQQ; 5.  
 CC PROSITE; PS00363; BACTERIAL\_PQQ\_1; 1.  
 CC PROSITE; PS00364; BACTERIAL\_PQQ\_2; 1.  
 CC Oxidoreductase; PQQ; Transmembrane; Periplasmic; Signal.  
 CC SIGNAL 1 33 POTENTIAL.  
 CC CHAIN 34 801  
 CC [PYRROLOQUINOLINE-QUINONE].  
 CC TRANSMEM 39 55  
 CC TRANSMEM 59 79 POTENTIAL.  
 CC TRANSMEM 94 108 POTENTIAL.  
 CC TRANSMEM 119 138 POTENTIAL.  
 CC ACT\_SITE 471 471 BASE (POTENTIAL).  
 CC SEQUENCE 801 AA; 86956 MW; 2F4EA222E5450D8 CRC64;  
 SQ  
 Query Match 39.9%; Score 43.5; DB 1; Length 801;  
 Best Local Similarity 47.1%; Pred. No. 63;  
 Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;  
 QY 3 DPAEEDKRFESHICR 19  
 Db 249 DPKLTKDSFQ-HLTCR 264  
 RESULT 22  
 ID GCTB\_ACIFE STANDARD; PRT; 265 AA.  
 AC Q59112;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glutamate CoA-transferase subunit B (EC 2.8.3.12) (GCT small  
 DE subunit).  
 DE GCTB.  
 OS Acidaminococcus fermentans.  
 OS Bacteria; Firmicutes; Clostridia; Clostridiales; Acidaminococcaceae;  
 OC Acidaminococcus.  
 OC NCBI\_taxid=905;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18 AND 43-65.  
 RC STRAIN=ATCC 25085;  
 RX MEDLINE=95045599; PubMed=7957258;  
 RA Mack M., Bendrat K., Zelder O., Eckel E., Linder D., Buckel W.;  
 RT "Location of the two genes encoding glutamate coenzyme  
 RT A-transferase at the beginning of the hydroxyglutarate operon in  
 RT Acidaminococcus fermentans."  
 RL Eur. J. Biochem. 226:41-51(1994).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.55 ANGSTROMS).  
 RC STRAIN=ATCC 25085;  
 RX MEDLINE=97238937; PubMed=9083111;

RA Jacob U., Mack M., Clausen T., Huber R., Buckel W., Messerschmidt A.;  
 RT "Glutamate CoA-transferase from Acidaminococcus fermentans: the  
 RL crystal structure reveals homology with other CoA-transferases";  
 CC Structure 5:415-426(1997).  
 CC -1- FUNCTION: CATALYZES THE TRANSFER OF THE COA MOIETY FROM ACETYL COA  
 CC TO (R)-2-HYDROXYGLUTARATE AND RELATED COMPOUNDS LIKE GLUTACONATE.  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + (E)-glutamate = acetate +  
 CC glutacoyl-1-CoA.  
 CC -1- PATHWAY: Glutamate fermentation.  
 CC -1- SUBUNIT: Heterooctamer of four A and four B subunits.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE GLUTACONATE COA-TRANSFERASE SUBUNIT B  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X81440; CAA57200.1; -.  
 CC PIR; S51052; S51052.  
 CC PDB; 1POI; 18-MAR-98.  
 CC Transferase; 3D-structure.  
 CC KW INIT MET 0 0  
 CC FT ACT\_SITE 53 53  
 CC FT HELIX 8 17  
 CC FT TURN 18 19  
 CC FT TURN 22 23  
 CC FT STRAND 25 27  
 CC FT HELIX 32 42  
 CC FT TURN 43 44  
 CC FT TURN 46 47  
 CC FT STRAND 49 52  
 CC FT TURN 53 55  
 CC FT STRAND 56 59  
 CC FT TURN 68 69  
 CC FT HELIX 71 74  
 CC FT TURN 75 75  
 CC FT STRAND 77 80  
 CC FT HELIX 83 96  
 CC FT TURN 97 97  
 CC FT STRAND 101 105  
 CC FT STRAND 109 110  
 CC FT TURN 112 113  
 CC FT STRAND 116 117  
 CC FT STRAND 120 122  
 CC FT STRAND 127 130  
 CC FT TURN 135 136  
 CC FT HELIX 137 143  
 CC FT STRAND 146 149  
 CC FT TURN 154 156  
 CC FT STRAND 157 157  
 CC FT STRAND 166 166  
 CC FT TURN 170 171  
 CC FT TURN 173 174  
 CC FT HELIX 175 178  
 CC FT TURN 179 180  
 CC FT TURN 183 184  
 CC FT STRAND 187 192  
 CC FT TURN 193 194  
 CC FT STRAND 195 199  
 CC FT TURN 201 203  
 CC FT STRAND 206 211  
 CC FT TURN 213 214  
 CC FT HELIX 217 222  
 CC FT TURN 223 223  
 CC FT TURN 231 232  
 CC FT STRAND 234 235  
 CC FT HELIX 241 249  
 CC FT TURN 250 251

FT TURN 253 254  
SQ SEQUENCE 265 AA; 29035 MW; 1E7FF61B42162FB4 CRC64;  
Query Match 39.4%; Score 43; DB 1; Length 265;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIEC 18  
: ||| : ||| :  
DB 62 VEVPSVGLRFRMAHCG 79  
RESULT 23  
SURE METMA STANDARD; PRT; 267 AA.  
AC Q8PX31;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Acid phosphatase sure (EC 3.1.3.2).  
GN SURE OR WM1391.  
OS Methanosarcina mazei (Methanosarcina frisia).  
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2209;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Goel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;  
RX MEDLINE=22120827; PubMed=12125824;  
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,  
RA Martinez-Arias R., Henne A., Wlezer A., Baumer S., Jacobi C.,  
RA Bruggemann H., Lienard T., Christmann A., Boencke M., Steckel S.,  
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
RA Fritz H.-J., Gottschalk G.;  
RT "The genome of Methanosarcina mazei: evidence for lateral gene  
transfer between Bacteria and Archaea."  
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
CC -|- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
alcohol + phosphate.  
CC -|- COFACTOR: Magnesium (By similarity).  
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -|- SIMILARITY: Belongs to the sure acid phosphatase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AE013373; AAM31087.1; -  
DR HAMAP; MF 00060; -; 1.  
DR InterPro; IPR002828; Sure.  
DR Pfam; PF01975; Sure; 1.  
DR ProDom; PD005378; Sure; 1.  
DR TIGRFAMs; TIGR00087; sure; 1.  
KW Hydrolase; Magnesium; Complete proteome.  
FT ACT\_SITE 133 133 POTENTIAL.  
FT METAL 14 14 MAGNESIUM (BY SIMILARITY).  
FT METAL 15 15 MAGNESIUM (BY SIMILARITY).  
FT METAL 45 45 MAGNESIUM (BY SIMILARITY).  
FT METAL 100 100 MAGNESIUM (BY SIMILARITY).  
SQ SEQUENCE 267 AA; 29097 MW; 709A83CC7854A771 CRC64;  
Query Match 39.4%; Score 43; DB 1; Length 267;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 2 DDPRESEEDKRFESHIE 17  
: ||| : ||| :  
DB 144 DDPDYQRFERFEAGIK 159  
RESULT 24  
RPBI\_EUPOC STANDARD; PRT; 478 AA.  
AC P28364;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6)  
DE (Fragment).  
DB RPBI.  
GN Euplates octocarinatus.  
OC Eukaryota; Alveolata; Ciliophora; Spirotrichea; Hypotrichia;  
OC Euplotida; Euplotidae; Euplates.  
OX NCBI\_TaxID=5937;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=(68)-VIII;  
RX MEDLINE=93027138; PubMed=1408746;  
RA Kaufmann J., Klein A.;  
RT "Gene dosage as a possible major determinant for equal expression  
levels of genes encoding RNA polymerase subunits in the hypotrichous  
ciliate Euplates octocarinatus";  
RL Nucleic Acids Res. 20:4445-4450(1992).  
CC -|- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
of DNA into RNA using the four ribonucleoside triphosphates as  
substrates.  
CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
(RNA)(N).  
CC -|- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM TEN TO  
TWELVE DIFFERENT POLYPEPTIDES. THIS POLYPEPTIDE IS THE LARGEST  
COMPONENT OF RNA POLYMERASE II.  
CC -|- SUBCELLULAR LOCATION: Nuclear.  
CC -|- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases are  
found in eukaryotic nuclei: polymerase I for the ribosomal RNA  
precursor, polymerase II for the mRNA precursor, and polymerase  
III for 5S and tRNA genes.  
CC -|- SIMILARITY: Belongs to the RNA polymerase beta' chain family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X66452; CAA47068.1; -  
DR PIR; S33886; S33886.  
DR InterPro; IPR000722; RNA\_pol\_A.  
DR InterPro; IPR007080; RNA\_pol\_Rpbi\_1.  
DR InterPro; IPR008592; RNA\_pol\_A\_N.  
DR Pfam; PF04997; RNA\_pol\_Rpbi\_1; 1.  
DR Pfam; PF00623; RNA\_pol\_Rpbi\_2; 1.  
DR SMART; SM00663; RPOLA\_N; 1.  
KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
DNA-binding; Nuclear protein; Zinc-finger.  
FT ZN\_FING 68 84 C2H2-TYPE (POTENTIAL).  
FT NON\_TER 478 478  
SQ SEQUENCE 478 AA; 54026 MW; 47A1ADD6ED98EB3D CRC64;  
Query Match 39.4%; Score 43; DB 1; Length 478;  
Best Local Similarity 47.4%; Pred. No. 43;  
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
QY 2 DDPRESEEDKRFESHIECK 20  
: ||| : ||| :  
DB 186 DDPSTNDKRDLSASECLK 204  
RESULT 25  
CSID\_ARATH STANDARD; PRT; 502 AA.  
ID CSID\_ARATH

FT TURN 253 254  
SQ SEQUENCE 265 AA; 29035 MW; 1E7FF61B42162FB4 CRC64;  
Query Match 39.4%; Score 43; DB 1; Length 265;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIEC 18  
: ||| : ||| :  
DB 62 VEVPSVGLRFRMAHCG 79  
RESULT 23  
SURE METMA STANDARD; PRT; 267 AA.  
AC Q8PX31;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Acid phosphatase sure (EC 3.1.3.2).  
GN SURE OR WM1391.  
OS Methanosarcina mazei (Methanosarcina frisia).  
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2209;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Goel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;  
RX MEDLINE=22120827; PubMed=12125824;  
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,  
RA Martinez-Arias R., Henne A., Wlezer A., Baumer S., Jacobi C.,  
RA Bruggemann H., Lienard T., Christmann A., Boencke M., Steckel S.,  
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
RA Fritz H.-J., Gottschalk G.;  
RT "The genome of Methanosarcina mazei: evidence for lateral gene  
transfer between Bacteria and Archaea."  
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
CC -|- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
alcohol + phosphate.  
CC -|- COFACTOR: Magnesium (By similarity).  
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -|- SIMILARITY: Belongs to the sure acid phosphatase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AE013373; AAM31087.1; -  
DR HAMAP; MF 00060; -; 1.  
DR InterPro; IPR002828; Sure.  
DR Pfam; PF01975; Sure; 1.  
DR ProDom; PD005378; Sure; 1.  
DR TIGRFAMs; TIGR00087; sure; 1.  
KW Hydrolase; Magnesium; Complete proteome.  
FT ACT\_SITE 133 133 POTENTIAL.  
FT METAL 14 14 MAGNESIUM (BY SIMILARITY).  
FT METAL 15 15 MAGNESIUM (BY SIMILARITY).  
FT METAL 45 45 MAGNESIUM (BY SIMILARITY).  
FT METAL 100 100 MAGNESIUM (BY SIMILARITY).  
SQ SEQUENCE 267 AA; 29097 MW; 709A83CC7854A771 CRC64;  
Query Match 39.4%; Score 43; DB 1; Length 267;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 2 DDPRESEEDKRFESHIE 17  
: ||| : ||| :  
DB 144 DDPDYQRFERFEAGIK 159

AC Q9FG65; 065789;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Cytochrome P450 81D1 (EC 1.14.-.-)  
 GN CYP81D1 OR CYP91A1 OR A75G36220 OR T30G6.3.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016721; PubMed=11130714;  
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,  
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,  
 RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,  
 RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,  
 RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,  
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,  
 RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,  
 RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,  
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,  
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,  
 RA Leonard S., Meyer R., Mulvaney E., Ozerky P., Riley A., Strommatt C.,  
 RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,  
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,  
 RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,  
 RA Martensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,  
 RA Volckaert G., Kambutt R., Duesterhoeft A., Stiekema W., Pohl T.,  
 RA Etian K.-D., Terry N., Hartley N., Bent E., Johnson S.,  
 RA Langham S.-A., McCullagh B., Robben J., Grymoprez B., Zimmermann W.,  
 RA Ransperger U., Wedler H., Balke K., Wedler E., Peters S.,  
 RA van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,  
 RA Weitzenecker T., Bothe G., Rose M., Hauf J., Bernsner S., Hempel S.,  
 RA Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,  
 RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,  
 RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.F.;  
 RA "Sequence and analysis of chromosome 5 of the plant Arabidopsis  
 RA thaliana."  
 RL Nature 408:823-826(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=22954850; PubMed=14593172;  
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
 RA Miranda M., Quach H.B., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
 RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,  
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
 RA Hayashizaki Y., Johnson-Hopson C., Heuan V.W., Iida K., Karnes M.,  
 RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
 RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
 RA Satou M., Tamse R., Vayberg M., Wallender E.K., Wong C., Yamamura Y.,  
 RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RA "Empirical analysis of transcriptional activity in the Arabidopsis  
 RA genome."  
 RL Science 302:842-846(2003).  
 RN [3]  
 RP SEQUENCE OF 3-502 FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=98281573; PubMed=9620263;  
 RA Mizutani M., Ward E., Ohta D.;  
 RT "Cytochrome P450 superfamily in Arabidopsis thaliana. Isolation of  
 RT cDNAs, differential expression, and RFLP mapping of multiple  
 RT cytochromes P450."  
 RL Plant Mol. Biol. 37:39-52(1998).  
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AB026661; BAB09361.1; -  
 DR EMBL; D78606; BAA28538.1; -  
 DR EMBL; AY093766; AAM10388.1; -  
 DR HSSP; P14779; 1JFZ.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PRO0385; P450.  
 KW Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.  
 FT TRANSMEM 6 26 POTENTIAL.  
 FT METAL 440 440 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 FT CONFLICT 121 121 T -> A (IN REF. 2).  
 SQ SEQUENCE 502 AA; 56723 MW; 1A348A8DD76A41B0 CRC64;  
 Query Match 39.4%; Score 43; DB 1; Length 502;  
 Best Local Similarity 47.4%; Pred. No. 46;  
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
 QY 2 DDPSEEDKRFESHIECKK 20  
 DB 411 DDPDSFKPERKEEEAQK 429  
 RESULT 26  
 DNAME FUSNN  
 ID DNAME FUSNN STANDARD; PRT; 607 AA.  
 AC Q8RH05;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chapterone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa  
 DE protein) (HSP70).  
 DE DNAME OR FN0116.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;  
 OC Fusobacterium.  
 OC NCBI\_TaxID=76856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 25586;  
 RX MEDLINE=21886394; PubMed=11899109;  
 RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
 RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,  
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
 RA Fongstein M., Kyripides N., Overbeek R.;  
 RA "Genome sequence and analysis of the oral bacterium Fusobacterium  
 RA nucleatum strain ATCC 25586."  
 RT J. Bacteriol. 184:2005-2018(2002).  
 CC -1- FUNCTION: Acts as a chaperone (By similarity).  
 CC -1- INDUCTION: By stress conditions e.g. heat shock (By similarity).  
 CC -1- SIMILARITY: Belongs to the heat shock protein 70 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; AB010525; AAL94325.1; -  
 DR HAMAP; MF 00332; -; 1.  
 DR InterPro; IPR001023; Hsp70.  
 DR Pfam; PF00012; HSP70; 1.  
 DR PRINTS; PRO0301; HEATSHOCK70.





OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).  
RX MEDLINE=90307707; PubMed=2195026;  
RA Winkelmann J.C., Chang J.G., Tse W.T., Scarpa A.L., Marchesi V.T.,  
RA Forget B.G.;  
RT "Full-length sequence of the cDNA for human erythroid beta-spectrin.";  
RL J. Biol. Chem. 265:11827-11832(1990).  
RN [2]  
RP SEQUENCE OF 1055-2137 FROM N.A. (ISOFORM 2).  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=91056094; PubMed=2243099;  
RA Winkelmann J.C., Costa F., Linzie B.L., Forget B.G.;  
RT "Beta spectrin in human skeletal muscle. Tissue-specific differential  
RT processing of 3' beta spectrin pre-mRNA generates a beta spectrin  
RT isoform with a unique carboxyl terminus.";  
RL J. Biol. Chem. 265:20449-20454(1990).  
RN [3]  
RP SEQUENCE OF 2002-2137 FROM N.A.  
RX MEDLINE=91332035; PubMed=1840591;  
RA Gallagher P.G., Tse W.T., Costa F., Scarpa A., Boivin P., Delaunay J.,  
RA Forget B.G.;  
RT "A splice site mutation of the beta-spectrin gene causing exon  
RT skipping in hereditary elliptocytosis associated with a truncated  
RT beta-spectrin chain.";  
RL J. Biol. Chem. 266:15154-15159(1991).  
RN [4]  
RP SEQUENCE OF 928-1755 FROM N.A.  
RX MEDLINE=91007291; PubMed=1976574;  
RA Yoon S.H., Kentros C.G., Prchal J.T.;  
RT "Identification of an unusual deletion within homologous repeats of  
RT human reticulocyte beta-spectrin and probable peptide polymorphism.";  
RL Gene 91:297-302(1990).  
RN [5]  
RP SEQUENCE OF 1334-1432 AND 1903-2137 FROM N.A.  
RX MEDLINE=88269838; PubMed=3390609;  
RA Winkelmann J.C., Leto T.L., Watkins P.C., Eddy R., Shows T.B.,  
RA Linnenbach A.J., Sahr K.E., Kathuria N., Marchesi V.T., Forget B.G.;  
RT "Molecular cloning of the cDNA for human erythrocyte beta-spectrin.";  
RL Blood 72:328-334(1988).  
RN [6]  
RP SEQUENCE OF 1209-1482 FROM N.A.  
RX MEDLINE=88041127; PubMed=3478706;  
RA Prchal J.T., Morley B.J., Yoon S.H., Coetzer T.L., Palek J.,  
RA Conboy J.G., Kan Y.W.;  
RT "Isolation and characterization of cDNA clones for human erythrocyte  
RT beta-spectrin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:7468-7472(1987).  
RN [7]  
RP DOMAINS  
RX MEDLINE=84295638; PubMed=6472478;  
RA Speicher D.W., Marchesi V.T.;  
RT "Erythrocyte spectrin is comprised of many homologous triple helical  
RT segments.";  
RL Nature 311:177-180(1984).  
RN [8]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97001215; PubMed=8844207;  
RA Maillet P., Aloisio N., Morle L., Delaunay J.;  
RT "Spectrin mutations in hereditary elliptocytosis and hereditary  
RT spherocytosis.";  
RL Hum. Mutat. 8:97-107(1996).  
RN [9]  
RP VARIANT HE CAGLIARI GLY-2018.  
RX MEDLINE=94043025; PubMed=8226774;  
RA Sahr K.E., Coetzer T.L., Moy L.S., Derick L.H., Chishti A.H.,  
RA Jarolim P., Lorenzo F., del Giudice E.M., Iolascon A., Gallanello R.,  
RA Cao A., Delaunay J., Liu S.-C., Palek J.;  
RT "Spectrin Cagliari: an Ala--Gly substitution in helix 1 of beta  
RT spectrin repeat 17 that severely disrupts the structure and self-  
RT association of the erythrocyte spectrin heterodimer.";

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).  
RX MEDLINE=90307707; PubMed=2195026;  
RA Winkelmann J.C., Chang J.G., Tse W.T., Scarpa A.L., Marchesi V.T.,  
RA Forget B.G.;  
RT "Full-length sequence of the cDNA for human erythroid beta-spectrin.";  
RL J. Biol. Chem. 265:11827-11832(1990).  
RN [2]  
RP SEQUENCE OF 1055-2137 FROM N.A. (ISOFORM 2).  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=91056094; PubMed=2243099;  
RA Winkelmann J.C., Costa F., Linzie B.L., Forget B.G.;  
RT "Beta spectrin in human skeletal muscle. Tissue-specific differential  
RT processing of 3' beta spectrin pre-mRNA generates a beta spectrin  
RT isoform with a unique carboxyl terminus.";  
RL J. Biol. Chem. 265:20449-20454(1990).  
RN [3]  
RP SEQUENCE OF 2002-2137 FROM N.A.  
RX MEDLINE=91332035; PubMed=1840591;  
RA Gallagher P.G., Tse W.T., Costa F., Scarpa A., Boivin P., Delaunay J.,  
RA Forget B.G.;  
RT "A splice site mutation of the beta-spectrin gene causing exon  
RT skipping in hereditary elliptocytosis associated with a truncated  
RT beta-spectrin chain.";  
RL J. Biol. Chem. 266:15154-15159(1991).  
RN [4]  
RP SEQUENCE OF 928-1755 FROM N.A.  
RX MEDLINE=91007291; PubMed=1976574;  
RA Yoon S.H., Kentros C.G., Prchal J.T.;  
RT "Identification of an unusual deletion within homologous repeats of  
RT human reticulocyte beta-spectrin and probable peptide polymorphism.";  
RL Gene 91:297-302(1990).  
RN [5]  
RP SEQUENCE OF 1334-1432 AND 1903-2137 FROM N.A.  
RX MEDLINE=88269838; PubMed=3390609;  
RA Winkelmann J.C., Leto T.L., Watkins P.C., Eddy R., Shows T.B.,  
RA Linnenbach A.J., Sahr K.E., Kathuria N., Marchesi V.T., Forget B.G.;  
RT "Molecular cloning of the cDNA for human erythrocyte beta-spectrin.";  
RL Blood 72:328-334(1988).  
RN [6]  
RP SEQUENCE OF 1209-1482 FROM N.A.  
RX MEDLINE=88041127; PubMed=3478706;  
RA Prchal J.T., Morley B.J., Yoon S.H., Coetzer T.L., Palek J.,  
RA Conboy J.G., Kan Y.W.;  
RT "Isolation and characterization of cDNA clones for human erythrocyte  
RT beta-spectrin.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:7468-7472(1987).  
RN [7]  
RP DOMAINS  
RX MEDLINE=84295638; PubMed=6472478;  
RA Speicher D.W., Marchesi V.T.;  
RT "Erythrocyte spectrin is comprised of many homologous triple helical  
RT segments.";  
RL Nature 311:177-180(1984).  
RN [8]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97001215; PubMed=8844207;  
RA Maillet P., Aloisio N., Morle L., Delaunay J.;  
RT "Spectrin mutations in hereditary elliptocytosis and hereditary  
RT spherocytosis.";  
RL Hum. Mutat. 8:97-107(1996).  
RN [9]  
RP VARIANT HE CAGLIARI GLY-2018.  
RX MEDLINE=94043025; PubMed=8226774;  
RA Sahr K.E., Coetzer T.L., Moy L.S., Derick L.H., Chishti A.H.,  
RA Jarolim P., Lorenzo F., del Giudice E.M., Iolascon A., Gallanello R.,  
RA Cao A., Delaunay J., Liu S.-C., Palek J.;  
RT "Spectrin Cagliari: an Ala--Gly substitution in helix 1 of beta  
RT spectrin repeat 17 that severely disrupts the structure and self-  
RT association of the erythrocyte spectrin heterodimer.";

RESULT 28

YQ36 CAEEL STANDARD; PRT; 963 AA.  
AC Q09457;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Putative cuticle collagen C09G5.6.  
GN C09G5.6  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Palmer S.;  
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Nematode cuticles are composed largely of collagen-like  
CC proteins. The cuticle functions both as an exoskeleton and as a  
CC barrier to protect the worm from its environment (By similarity).  
CC -!- SUBUNIT: Collagen polypeptide chains are complexed within the  
CC cuticle by disulfide bonds and other types of covalent cross-links  
CC (By similarity).  
CC -!- SIMILARITY: Belongs to the cuticular collagen family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z46791; CAA86755.1; --  
CC PIR; T19140; T19140.  
CC WORMPEP; C09G5.6; CE01486.  
CC InterPro; IPR008161; C1G\_helix.  
CC InterPro; IPR002486; Col\_cuticle\_N.  
CC InterPro; IPR008160; Collagen.  
CC Pfam; PF01484; Col\_cuticle\_N; 1.  
CC Pfam; PF01391; Collagen; 3.  
CC ProDom; PD000007; C1G\_helix; 2.  
CC Hypothetical protein; Cuticle; Connective tissue; Repeat;  
CC Multigene family; Collagen.  
CC DOMAIN 392; 423 TRIPLE-HELICAL REGION.  
FT DOMAIN 441; 503 TRIPLE-HELICAL REGION.  
FT DOMAIN 506; 567 TRIPLE-HELICAL REGION.  
FT DOMAIN 563; 666 POLY-PRO.  
FT DOMAIN 685; 688 POLY-PRO.  
SQ SEQUENCE 963 AA; 107031 MW; AFF895A75909F66E CRC64;

Query Match 39.4%; Score 43; DB 1; Length 963;  
Best Local Similarity 50.0%; Pred. No. 92;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 PRSEEDKRFESHIECR 19

Db 777 PQEPRRYTEAPR 792

RESULT 29

SPCB HUMAN  
ID SPCB HUMAN STANDARD; PRT; 2137 AA.  
AC P1277; Q15510; Q15519;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Spectrin beta chain, erythrocyte (Beta-I spectrin).  
GN SPTB OR SPTBI.  
OS Homo sapiens (Human).

RL J. Biol. Chem. 268:22656-22662(1993).  
RN [10]  
RP VARIANT HS KISSIMMEE ARG-202.  
RX MEDLINE=93352802; PubMed=8102379;  
RA Becker P.S., Tse W.T., Lux S.E., Forget B.G.;  
RT "Beta spectrin KISSimmee: a spectrin variant associated with  
RT autosomal dominant hereditary spherocytosis and defective binding to  
RT protein 4.1";  
RL J. Clin. Invest. 92:612-616(1993).  
RN [11]  
RP VARIANT HE PROVIDENCE PRO-2019.  
RX MEDLINE=95190014; PubMed=7883966;  
RA Gallagher P.G., Weed S.A., Tse W.T., Benoit L., Morrow J.S.,  
RA Marchesi S.L., Mohandas N., Forget B.G.;  
RT "Recurrent fatal hydrops fetalis associated with a nucleotide  
RT substitution in the erythrocyte beta-spectrin gene";  
RL J. Clin. Invest. 95:1174-1182(1995).  
RN [12]  
RP VARIANTS HE VAL-2023 AND ARG-2024.  
RX MEDLINE=94289716; PubMed=8018926;  
RA Parquet N., Devaux I., Boulanger L., Galand C., Boivin P.,  
RA Lecomte M.-C., Dhery D., Garbarz M.;  
RT "Identification of three novel spectrin alpha I/74 mutations in  
RT hereditary elliptocytosis: further support for a triple-stranded  
RT folding unit model of the spectrin heterodimer contact site";  
RL Blood 84:303-308(1994).  
RN [13]  
RP VARIANT HE PRO-2053.  
RX MEDLINE=90369011; PubMed=1375598;  
RA Tse W.T., Lecomte M.-C., Costa F.F., Garbarz M., Feo C., Boivin P.,  
RA Dhery D., Forget B.G.;  
RT "Point mutation in the beta-spectrin gene associated with alpha I/74  
RT hereditary elliptocytosis. Implications for the mechanism of spectrin  
RT dimer self-association";  
RL J. Clin. Invest. 86:909-916(1990).  
CC -!- FUNCTION: Spectrin is the major constituent of the cytoskeletal  
CC network underlying the erythrocyte plasma membrane. It associates  
CC with band 4.1 and actin to form the cytoskeletal superstructure of  
CC the erythrocyte plasma membrane.  
CC -!- SUBUNIT: Composed of nonhomologous chains, alpha and beta, which  
CC aggregate to form dimers, tetramers, and higher polymers.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=1;  
CC IsoId=P11277-1; Sequences=Displayed;  
CC Name=2; Synonyms=Muscle-specific;  
CC IsoId=P11277-2; Sequences=VSP\_000719;  
CC Name=3;  
CC IsoId=P11277-3; Sequences=VSP\_007242;  
CC -!- DISEASE: Defects in SPTB are a cause of rhesus-unlinked hereditary  
CC elliptocytosis (HE) [MIM:182870, 130600, 166900]; also called  
CC elliptocytosis 2 (EL2). Hereditary elliptocytosis (also known as  
CC hereditary ovalocytosis) is a genetically heterogeneous, autosomal  
CC dominant hemolytic disorder. It is characterized by variable  
CC hemolytic anemia and elliptical or oval red cell shape. Several  
CC abnormalities in the membrane skeleton have been identified,  
CC including a number that have been localized to spectrin.  
CC Elliptocytosis due to defects in SPTB can also be referred to as  
CC elliptocytosis 3 (EL3).  
CC -!- DISEASE: Defects in SPTB are the cause of spherocytosis type I  
CC (SPH1) [MIM:182870], a disorder characterized by severe hemolytic  
CC anemia. Inheritance is autosomal dominant.  
CC -!- MISCELLANEOUS: This complex is anchored to the cytoplasmic face of  
CC the plasma membrane via another protein, ankyrin, which binds to  
CC beta-spectrin and mediates the binding of the whole complex to a  
CC transmembrane protein band 3. The interaction of erythrocyte  
CC spectrin with other proteins through specific binding domains lead  
CC to the formation of an extensive subplasmalemmal meshwork which is  
CC thought to be responsible for the maintenance of the biconcave  
CC shape of human erythrocytes, for the regulation of plasma membrane  
CC components and for the maintenance of the lipid asymmetry of the  
CC plasma membrane.  
CC -!- SIMILARITY: Belongs to the spectrin family.

CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
CC -!- SIMILARITY: Contains 17 spectrin repeats.  
CC -----  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: J05500; AAA60578.1; --  
CC EMBL: J05500; AAA60579.1; --  
CC EMBL: M37884; AAA63289.1; --  
CC EMBL: M37885; AAA60571.1; --  
CC EMBL: M57948; --; NOT ANNOTATED\_CDS.  
CC EMBL: X59510; CAA42037.1; --  
CC EMBL: X59511; CAA42036.1; --  
CC EMBL: M18054; AAA60572.1; --  
CC PIR: A37064; SJHJB.  
CC HSP: O01082; 1BKR.  
CC SWISS-2DPAGE: P11277; HUMAN.  
CC Genew: HGNC:11274; SPTB.  
CC MIM: 182870; --  
CC MIM: 130600; --  
CC MIM: 166900; --  
CC GO: GO:0008091; C-spectrin; TAS.  
CC GO: GO:0003779; F-actin binding; TAS.  
CC InterPro: IPR001589; Actbind actinin.  
CC InterPro: IPR001715; Calponin-like.  
CC InterPro: IPR002017; Spectrin.  
CC Pfam: PF00307; CH; 2.  
CC Pfam: PF00435; spectrin; 17.  
CC SMART: SM00033; CH; 2.  
CC PROSITE: PS000150; SPEC; 17.  
CC PROSITE: PS00019; ACTININ 1; 1.  
CC PROSITE: PS00020; ACTININ 2; 1.  
CC PROSITE: PS50021; CH; 2.  
CC Cytoskeleton; Membrane; Erythrocyte; Repeat; Actin-binding;  
CC Disease mutation; Hereditary hemolytic anemia; Elliptocytosis;  
CC Alternative splicing; Actin capping.  
CC DOMAIN 1 275 ACTIN-BINDING.  
CC FT DOMAIN 54 158 CH 1.  
CC FT DOMAIN 173 275 CH 2.  
CC FT REPEAT 276 384 SPECTRIN 1.  
CC FT REPEAT 385 498 SPECTRIN 2.  
Query Match 39.4%; Score 43; DB 1; Length 2137;  
Best Local Similarity 43.8%; Pred. No. 2.2e+02;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
Qy 4 PRSEEDKRFESHIECR 19  
Db 920 PRSEVKQYQDHLNTR 935  
|||||:::  
RESULT 30  
SFAS CHLEU STANDARD; PRT; 269 AA.  
ID SFAS CHLEU  
AC Q39430;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DB SF-assemblin.  
OS Chlamydomonas eugametos.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3053;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97167588; PubMed=9015206;  
RA Lechtreck K.-F., Silflow C.;  
RT "SF-assemblin in Chlamydomonas: sequence conservation and localization

```
RT during the cell cycle."
RL Cell Motil. Cytoskeleton 36:190-201(1997).
CC -!- FUNCTION: Major component of the striated microtubule-associated
CC fibers (SMAFs; system-I-fibers).
CC -!- DOMAIN: Consists of a small non-helical N-terminal domain and a
CC rod domain with a 29 residue repeat pattern based on four heptads
CC followed by a skip residue. This alpha-helical protein is
CC characterized by the ability to form a special segmented coiled
CC coil and to assemble into striated fibers of 2 nm protofilaments
CC (By similarity).
CC -!- SIMILARITY: TO OTHER GREEN ALGAE SF-ASSEMBLIN AND GIARDIA
CC LAMBLLA BETA-GIARDIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; U56981; AAB03788.1; -.
CC InterPro: IPR008374; SF assemblin.
CC PRINTS; PR01799; SFASSEMBLIN.
CC KW Cytoskeleton; Microtubule; Coiled coil.
CC FT DOMAIN 1 24 NONHELICAL REGION.
CC FT DOMAIN 25 269 ROD.
CC FT DOMAIN 98 144 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 269 AA; 30444 MW; 38CE792501C307D0 CRC64;
CC
CC Query Match 38.5%; Score 42; DB 1; Length 269;
CC Best Local Similarity 53.8%; Pred. No. 33;
CC Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
CC
CC Qy 5 RSEEDKRPESHIE 17
CC Db 79 RAESDKQLQSHFE 91
```

Search completed: May 11, 2004, 13:37:58  
Job time : 8.75 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:28:26 ; Search time 32 Seconds  
(without alignments)  
197.199 Million cell updates/sec

Title: US-09-171-432A-38

Perfect score: 109  
Sequence: 1 VDDPSHEDKRFESHIECRK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	56	12	Q67819 hepatitis a
2	109	100.0	56	12	Q67818 hepatitis a
3	109	100.0	56	12	Q67820 hepatitis a
4	109	100.0	56	12	Q39865 hepatitis a
5	109	100.0	56	12	Q7ICYS hepatitis a
6	109	100.0	56	12	Q8JMR7 hepatitis a
7	109	100.0	56	12	Q8JMS2 hepatitis a
8	109	100.0	56	12	Q8JN35 hepatitis a
9	109	100.0	56	12	Q8JN35 hepatitis a
10	109	100.0	56	12	Q8JN35 hepatitis a
11	109	100.0	56	12	Q8JN35 hepatitis a
12	109	100.0	56	12	Q8JN35 hepatitis a
13	109	100.0	56	12	Q8JN35 hepatitis a
14	109	100.0	56	12	Q8JN35 hepatitis a
15	109	100.0	56	12	Q8JN35 hepatitis a
16	109	100.0	56	12	Q8JN35 hepatitis a
17	109	100.0	56	12	Q8JN35 hepatitis a

90 106 97.2 56 12 Q8JYK3 hepatitis a  
 91 106 97.2 56 12 Q9IPD0 hepatitis a  
 92 106 97.2 56 12 Q9IPB3 hepatitis a  
 93 106 97.2 56 12 Q9IPB2 hepatitis a  
 94 106 97.2 56 12 Q99U7 hepatitis a  
 95 106 97.2 56 12 Q99H1 hepatitis a  
 96 106 97.2 56 12 Q99H1 hepatitis a  
 97 106 97.2 56 12 Q99H1 hepatitis a  
 98 106 97.2 56 12 Q99H1 hepatitis a  
 99 106 97.2 56 12 Q8JYK2 hepatitis a  
 100 106 97.2 56 12 Q99U6 hepatitis a

ALIGNMENTS

RESULT 1  
 Q67822 PRELIMINARY; PRT; 56 AA.  
 AC Q67822;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ISR-70;  
 RA Karetvi Y.V., Shulman L.M., Manor J., Leitner L., Shehab S.,  
 RA Mendelson E.;  
 RT "Genetic classification of hepatitis A virus strains isolated in  
 RT Israel, based on their VP1/2A nucleotide sequence.";  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z77247; CAB01040.1; -;  
 DR PIR; Q0427; Q0427.  
 DR PIR; Q0428; Q0428.  
 DR PIR; Q0431; Q0431.  
 DR InterPro; IPR000886; ER target\_S.  
 DR PROSITE; PS00014; ER TARGET; 1.  
 FT NON TER 1  
 FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
 FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
 FT NON TER 56  
 SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEFA44 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
 DB 16 VDDPRSEEDKRFESHIECKR 35

RESULT 2  
 Q39867 PRELIMINARY; PRT; 56 AA.  
 AC Q39867;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JVR;  
 RX MEDLINE=97247817; PubMed=9093940;

RA Taylor M.B.;  
 RT "Molecular epidemiology of South African strains of hepatitis A virus:  
 RT 1982-1996.";  
 RL J. Med. Virol. 51:273-279(1997).  
 DR EMBL; U68692; AAB53588.1; -;  
 DR PIR; Q0427; Q0427.  
 DR PIR; Q0428; Q0428.  
 DR PIR; Q0431; Q0431.  
 DR InterPro; IPR000886; ER target\_S.  
 DR PROSITE; PS00014; ER TARGET; 1.  
 FT NON TER 1  
 FT NON TER 56  
 SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEFA44 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
 DB 16 VDDPRSEEDKRFESHIECKR 35

RESULT 3  
 Q918Q3 PRELIMINARY; PRT; 56 AA.  
 AC Q918Q3;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RJ-005;  
 RA de Paula V.S., Baptista M.L., Lampe E., Niel C., Gaspar A.M.C.;  
 RT "Characterization of hepatitis A virus isolates from subgenotypes IA  
 RT and IB in Rio de Janeiro, Brazil.";  
 RL J. Med. Virol. 0:0-0(2001).  
 DR EMBL; AF410381; AAL0177.1; -;  
 DR PIR; Q0427; Q0427.  
 DR PIR; Q0428; Q0428.  
 DR PIR; Q0431; Q0431.  
 DR InterPro; IPR000886; ER target\_S.  
 DR PROSITE; PS00014; ER TARGET; 1.  
 FT NON TER 1  
 FT NON TER 56  
 SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEFA44 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
 DB 16 VDDPRSEEDKRFESHIECKR 35

RESULT 4  
 Q91P95 PRELIMINARY; PRT; 56 AA.  
 AC Q91P95;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;

RN [1] SEQUENCE FROM N.A.  
RP STRAIN=A7;  
RC FUJIWARA K.;  
RX "hepatitis A virus VP1/2A junction."  
RA Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RT EMBL; AB046915; BAB08072.1; -  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;  
  
Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
Db 16 VDDPRSEEDKRFESHIECRK 35  
  
RESULT 5  
Q918Q0 PRELIMINARY; PRT; 56 AA.  
ID Q918Q0;  
AC Q918Q0;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polyprotein (fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RJ-055;  
RA de Paula V.S., Baptista M.L., Lampe E., Niel C., Gaspar A.M.C.;  
RT "Characterization of hepatitis A virus isolates from subgenotypes IA and IB in Rio de Janeiro, Brazil."  
RL J. Med. Virol. 0:0-0(2001).  
DR EMBL; AF410384; AAL10180.1; -  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6598 MW; E838C51846AEF4B4 CRC64;  
  
Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
Db 16 VDDPRSEEDKRFESHIECRK 35  
  
RESULT 6  
Q918Q0 PRELIMINARY; PRT; 56 AA.  
ID Q918Q0;  
AC Q918Q0;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polyprotein (fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;

RN [1] SEQUENCE FROM N.A.  
RP STRAIN=96001190;  
RX MEDLINE=97247817; PubMed=9093940;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus: 1982-1996."  
RL J. Med. Virol. 51:273-279(1997).  
DR EMBL; U68690; AAB53586.1; -  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;  
  
Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
Db 16 VDDPRSEEDKRFESHIECRK 35  
  
RESULT 7  
Q67821 PRELIMINARY; PRT; 56 AA.  
ID Q67821;  
AC Q67821;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polyprotein (fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISR-15;  
RA Karetnyi Y.V., Shulman L.M., Manor J., Leitner L., Shehab S., Mendelson E.;  
RT "Genetic classification of hepatitis A virus strains isolated in Israel, based on their VP1/2A nucleotide sequence."  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z77246; CAB01039.1; -  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;  
  
Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
Db 16 VDDPRSEEDKRFESHIECRK 35  
  
RESULT 8  
Q918Q4 PRELIMINARY; PRT; 56 AA.  
ID Q918Q4  
AC Q918Q4;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)

01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
Polypeptide (Fragment)  
Hepatitis A virus.  
Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
Hepatovirus.  
NCBI\_TaxID=12092;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=RU-004;  
de Paula V.S., Baptista M.L., Lampe E., Niel C., Gaspar A.M.C.;  
"Characterization of Hepatitis A virus isolates from subgenotypes IA  
and IB in Rio de Janeiro, Brazil.";  
J. Med. Virol. 0:0-0(2001).  
EMBL; AF410380; AAL10176.1; -  
PIR; PQ0427; PQ0428.  
PIR; PQ0428; PQ0428.  
PIR; PQ0431; PQ0431.  
InterPro; IPR000886; ER target S.  
PROSITE; PS00014; ER\_TARGET; 1.  
NON\_TER 1  
NON\_TER 56  
SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;  
Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIECK 20  
DB 16 VDDPRSEEDKRFESHIECK 35  
RESULT 9  
Q91PA8 PRELIMINARY; PRT; 56 AA.  
AC Q91PA8;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
DE Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=A40;  
Fujiwara K.;  
"Hepatitis A virus VP1/2A junction.";  
Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
EMBL; AB046902; BAB08059.1; -  
PIR; PQ0427; PQ0428.  
PIR; PQ0428; PQ0428.  
PIR; PQ0431; PQ0431.  
InterPro; IPR000886; ER target S.  
PROSITE; PS00014; ER\_TARGET; 1.  
NON\_TER 1  
NON\_TER 56  
SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;  
Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIECK 20  
DB 16 VDDPRSEEDKRFESHIECK 35  
RESULT 10  
Q67823 PRELIMINARY; PRT; 56 AA.  
ID Q67823

Q67823;  
01-NOV-1996 (TrEMBLrel. 01, Created)  
01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
DE Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=ISR-79;  
Karetnyi Y.V., Shulman L.M., Manor J., Leitner L., Shehab S.,  
Mendelson E.;  
"Genetic classification of hepatitis A virus strains isolated in  
Israel, based on their VP1/2A nucleotide sequence.";  
Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
EMBL; Z77248; CAB01041.1; -  
PIR; PQ0427; PQ0427.  
PIR; PQ0428; PQ0428.  
PIR; PQ0431; PQ0431.  
InterPro; IPR000886; ER target S.  
PROSITE; PS00014; ER\_TARGET; 1.  
NON\_TER 1  
NON\_TER 56  
CAPSID PROTEIN VP1.  
CHAIN 29 >56  
NONSTRUCTURAL PROTEIN NS2A.  
NON\_TER 56  
SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;  
Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIECK 20  
DB 16 VDDPRSEEDKRFESHIECK 35  
RESULT 11  
Q39874 PRELIMINARY; PRT; 56 AA.  
ID Q39874;  
AC Q39874;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
DE Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
NCBI\_TaxID=12092;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=314274;  
RC STRAIN=97247817; PubMed=9093940;  
Taylor M.B.;  
"Molecular epidemiology of South African strains of hepatitis A virus:  
1982-1996.";  
J. Med. Virol. 51:273-279(1997).  
EMBL; U68699; AAB53595.1; -  
PIR; PQ0427; PQ0427.  
PIR; PQ0428; PQ0428.  
PIR; PQ0431; PQ0431.  
InterPro; IPR000886; ER target S.  
PROSITE; PS00014; ER\_TARGET; 1.  
NON\_TER 1  
NON\_TER 56  
SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;  
Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIECK 20

1 VDDPRSEEDKRFESHIECRK 20  
16 VDDPRSEEDKRFESHIECRK 35

QY Db

RESULT 14

Q39868  
ID Q39868 PRELIMINARY; PRT; 56 AA.  
AC Q39868;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
DE Hepatitis A virus.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RC STRAIN=97247817; PubMed=9093940;  
RA de Paula V.S.; Baptista M.L., Lampe E., Niel C., Gaspar A.M.C.;  
RT "Characterization of hepatitis A virus isolates from subgenotypes IA  
and IB in Rio de Janeiro, Brazil."  
J. Med. Virol. 0:0-0(2001).  
RL J. Med. Virol. 51:273-279(1997).  
RL EMBL; U68693; AAB53590.1; -.  
DR PIR; PQ0427; PQ0428.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY Db

RESULT 15

Q39869  
ID Q39869 PRELIMINARY; PRT; 56 AA.  
AC Q39869;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
DE Hepatitis A virus.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RC STRAIN=96002382;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus;  
1982-1996."  
J. Med. Virol. 51:273-279(1997).  
RL EMBL; U68694; AAB53590.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY Db

RESULT 13

Q39869  
ID Q39869 PRELIMINARY; PRT; 56 AA.  
AC Q39869;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
DE Hepatitis A virus.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RC STRAIN=96002382;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus;  
1982-1996."  
J. Med. Virol. 51:273-279(1997).  
RL EMBL; U68694; AAB53590.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY Db

RESULT 12

Q39869  
ID Q39869 PRELIMINARY; PRT; 56 AA.  
AC Q39869;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
DE Hepatitis A virus.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RC STRAIN=97247817; PubMed=9093940;  
RA de Paula V.S.; Baptista M.L., Lampe E., Niel C., Gaspar A.M.C.;  
RT "Characterization of hepatitis A virus isolates from subgenotypes IA  
and IB in Rio de Janeiro, Brazil."  
J. Med. Virol. 0:0-0(2001).  
RL J. Med. Virol. 51:273-279(1997).  
RL EMBL; U68693; AAB53589.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6598 MW; 838C51846AEF4B4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY Db

RESULT 13

Q39869  
ID Q39869 PRELIMINARY; PRT; 56 AA.  
AC Q39869;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
DE Hepatitis A virus.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RC STRAIN=96002382;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus;  
1982-1996."  
J. Med. Virol. 51:273-279(1997).  
RL EMBL; U68694; AAB53590.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY Db

RESULT 12

Q39869  
ID Q39869 PRELIMINARY; PRT; 56 AA.  
AC Q39869;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
DE Hepatitis A virus.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RC STRAIN=97247817; PubMed=9093940;  
RA de Paula V.S.; Baptista M.L., Lampe E., Niel C., Gaspar A.M.C.;  
RT "Characterization of hepatitis A virus isolates from subgenotypes IA  
and IB in Rio de Janeiro, Brazil."  
J. Med. Virol. 0:0-0(2001).  
RL J. Med. Virol. 51:273-279(1997).  
RL EMBL; U68693; AAB53589.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6598 MW; 838C51846AEF4B4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY Db

RESULT 13

Q39869  
ID Q39869 PRELIMINARY; PRT; 56 AA.  
AC Q39869;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
DE Hepatitis A virus.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RC STRAIN=96002382;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus;  
1982-1996."  
J. Med. Virol. 51:273-279(1997).  
RL EMBL; U68694; AAB53590.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY Db

RESULT 12

Q39869  
ID Q39869 PRELIMINARY; PRT; 56 AA.  
AC Q39869;  
DT 01-DEC



Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20  
|||||  
DB 16 VDDPRSEEDKRFESHIECRK 35

RESULT 16

O39873 PRELIMINARY; PRT; 56 AA.

ID O39873;  
AC O39873;  
DT 01-JAN-1998 (TREMELrel. 05, Created)  
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Polypotein (Fragment).  
OS Hepatitis A virus.  
OC Hepatovirus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2333;  
RX MEDLINE=97247817; PubMed=9093940;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus:  
1982-1996.";  
RL J. Med. Virol. 51:273-279(1997).  
DR EMBL; U68698; AAB53594.1; -.  
DR PIR; PQ0427; PQ0428.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON TER 1 56  
FT NON TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20  
|||||  
DB 16 VDDPRSEEDKRFESHIECRK 35

RESULT 17

O67819 PRELIMINARY; PRT; 56 AA.

ID O67819;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Polypotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OX Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISR-10;  
RA Karetnyi Y.V.; Shulman L.M.; Manor J.; Leitner L.; Shehab S.;  
RA Mendelson E.;  
RT "Genetic classification of hepatitis A virus strains isolated in  
Israel, based on their VP1/2A nucleotide sequence.";  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z77244; CAB01037.1; -.  
DR PIR; PQ0427; PQ0428.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON TER 1 56  
FT NON TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20  
|||||  
DB 16 VDDPRSEEDKRFESHIECRK 35

RESULT 18

O67818 PRELIMINARY; PRT; 56 AA.

ID O67818;  
AC O67818;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Polypotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OX Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISR-05;  
RA Karetnyi Y.V.; Shulman L.M.; Manor J.; Leitner L.; Shehab S.;  
RA Mendelson E.;  
RT "Genetic classification of hepatitis A virus strains isolated in  
Israel, based on their VP1/2A nucleotide sequence.";  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z77243; CAB01036.1; -.  
DR PIR; PQ0427; PQ0428.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON TER 1 56  
FT NON TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20  
|||||  
DB 16 VDDPRSEEDKRFESHIECRK 35

RESULT 19

O67820 PRELIMINARY; PRT; 56 AA.

ID O67820;  
AC O67820;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Polypotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OX Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISR-12;  
RA Karetnyi Y.V.; Shulman L.M.; Manor J.; Leitner L.; Shehab S.;  
RA Mendelson E.;

FT NON TER 1 1  
FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
FT NON TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20  
|||||  
DB 16 VDDPRSEEDKRFESHIECRK 35

RESULT 18

O67818 PRELIMINARY; PRT; 56 AA.

ID O67818;  
AC O67818;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Polypotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OX Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISR-05;  
RA Karetnyi Y.V.; Shulman L.M.; Manor J.; Leitner L.; Shehab S.;  
RA Mendelson E.;  
RT "Genetic classification of hepatitis A virus strains isolated in  
Israel, based on their VP1/2A nucleotide sequence.";  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z77243; CAB01036.1; -.  
DR PIR; PQ0427; PQ0428.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON TER 1 56  
FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
FT NON TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECRK 20  
|||||  
DB 16 VDDPRSEEDKRFESHIECRK 35

RESULT 19

O67820 PRELIMINARY; PRT; 56 AA.

ID O67820;  
AC O67820;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Polypotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OX Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISR-12;  
RA Karetnyi Y.V.; Shulman L.M.; Manor J.; Leitner L.; Shehab S.;  
RA Mendelson E.;

RT "Genetic classification of hepatitis A virus strains isolated in  
RT Israel, based on their VP1/2A nucleotide sequence."  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z77245; CAB01038.1; -.  
DR PIR; P0427; P0427.  
DR PIR; P0428; P0428.  
DR PIR; P0431; P0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON\_TER 1 -1  
FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
|||  
DB 16 VDDPRSEEDKRFESHIECKR 35

RESULT 20

ID O39865 PRELIMINARY; PRT; 56 AA.  
AC O39865;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1] SEQUENCE FROM N.A.  
RC STRAIN=412991;  
RX MEDLINE=97247817; PubMed=9093940;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus:  
RT 1982-1996."  
RL J. Med. Virol. 51:273-279(1997).  
DR EMBL; U68689; AAB53585.1; -.  
DR PIR; P0427; P0427.  
DR PIR; P0428; P0428.  
DR PIR; P0431; P0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON\_TER 1 -1  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
|||  
DB 16 VDDPRSEEDKRFESHIECKR 35

RESULT 21

ID Q7TCY5 PRELIMINARY; PRT; 56 AA.  
AC Q7TCY5;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1] SEQUENCE FROM N.A.  
RC STRAIN=Arg9.6;  
RA Munne M.S., Vladimirovsky S.N., Gonzalez J.E.;  
RT "Molecular characterization of hepatitis A virus from Argentina."  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY314885; AAP79464.1; -.  
DR Polyprotein.  
FT NON\_TER 1 1  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
|||  
DB 16 VDDPRSEEDKRFESHIECKR 35

RESULT 22

ID Q8JMR7 PRELIMINARY; PRT; 59 AA.  
AC Q8JMR7;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1] SEQUENCE FROM N.A.  
RC STRAIN=A'Gam085;  
RA ter Bruggen J.P., Tjon G.M.S., Koek A.G., Wijkman C.J.,  
RA van Lieshout I.J., Leenders A.C.A.P., Bruisten S.M.;  
RT "Direct evidence of hepatitis A virus involvement in meningitis."  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY101271; AAM50074.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 59 59  
SQ SEQUENCE 59 AA; 6792 MW; FF358FBCD4F89D7 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 59;  
Best Local Similarity 100.0%; Pred. No. 6.1e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
|||  
DB 23 VDDPRSEEDKRFESHIECKR 42

RESULT 23

ID Q8JMS2 PRELIMINARY; PRT; 59 AA.  
AC Q8JMS2;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1] SEQUENCE FROM N.A.  
RC STRAIN=Brab008;  
RA ter Bruggen J.P., Tjon G.M.S., Koek A.G., Wijkman C.J.,  
RA van Lieshout I.J., Leenders A.C.A.P., Bruisten S.M.;

RT "Direct evidence of hepatitis A virus involvement in meningitis.";  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY101266; AAM50069.1; -  
FT NON\_TER 1 59  
FT NON\_TER 59 59  
SQ SEQUENCE 59 AA; 6820 MW; ED8EE8FBCD4F9D7 CRC64;  
  
Query Match 100.0%; Score 109; DB 12; Length 59;  
Best Local Similarity 100.0%; Pred. No. 6.1e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 23 VDDPRSEEDKRFESHIECRK 42  
|||||  
  
RESULT 24  
ID Q8BC97 PRELIMINARY; PRT; 71 AA.  
AC Q8BC97;  
DT 01-OCT-2002 (TremBLrel. 22, Created)  
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HAP300;  
RA De Paula V.S., Saback F.L., Gaspar A.M.C., Niel C.;  
RT "Mixed Infection of a Young Female Child Care Provider With Hepatitis  
RT A Virus Isolates From Subgenotypes IA and IB Revealed by Heteroduplex  
RT Mobility Assay.";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY083511; AAM03320.1; -  
FT NON\_TER 1 71  
FT NON\_TER 71 71  
SQ SEQUENCE 71 AA; 8322 MW; 05D9E23399F3E448 CRC64;  
  
Query Match 100.0%; Score 109; DB 12; Length 71;  
Best Local Similarity 100.0%; Pred. No. 7.4e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 32 VDDPRSEEDKRFESHIECRK 51  
|||||  
  
RESULT 25  
ID Q8BC97 PRELIMINARY; PRT; 72 AA.  
AC Q8BC97;  
DT 01-MAR-2003 (TremBLrel. 23, Created)  
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HAP163;  
RA Villar L.M., Lampe E., Meyer A., Gaspar A.M.C.;  
RT "Genetic variability of hepatitis A virus strains isolated during an  
RT outbreak in a public school in Rio de Janeiro, Brazil.";  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF538726; AAN15169.1; -  
FT NON\_TER 1 72  
FT NON\_TER 72 72  
SQ SEQUENCE 72 AA; 8425 MW; 06C6DAE2309AF3E4 CRC64;  
  
Query Match 100.0%; Score 109; DB 12; Length 72;  
Best Local Similarity 100.0%; Pred. No. 7.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 32 VDDPRSEEDKRFESHIECRK 51  
|||||  
  
RESULT 26  
ID Q8BC96 PRELIMINARY; PRT; 72 AA.  
AC Q8BC96;  
DT 01-MAR-2003 (TremBLrel. 23, Created)  
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HAP300;  
RA Villar L.M., Lampe E., Meyer A., Gaspar A.M.C.;  
RT "Genetic variability of hepatitis A virus strains isolated during an  
RT outbreak in a public school in Rio de Janeiro, Brazil.";  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF538727; AAN15169.1; -  
FT NON\_TER 1 72  
FT NON\_TER 72 72  
SQ SEQUENCE 72 AA; 8425 MW; 06C6DAE2309AF3E4 CRC64;  
  
Query Match 100.0%; Score 109; DB 12; Length 72;  
Best Local Similarity 100.0%; Pred. No. 7.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 32 VDDPRSEEDKRFESHIECRK 51  
|||||  
  
RESULT 27  
ID Q9QCT0 PRELIMINARY; PRT; 97 AA.  
AC Q9QCT0;  
DT 01-MAY-2000 (TremBLrel. 13, Created)  
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Capsid protein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MIB4022;  
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;  
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF199046; AAF08937.1; -  
FT NON\_TER 1 97  
FT NON\_TER 97 97  
SQ SEQUENCE 97 AA; 11282 MW; 5ECB8F22812A62B9 CRC64;

FT NON\_TER 72 72  
SQ SEQUENCE 72 AA; 8435 MW; 06C5D9E23399F3E4 CRC64;  
  
Query Match 100.0%; Score 109; DB 12; Length 72;  
Best Local Similarity 100.0%; Pred. No. 7.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 32 VDDPRSEEDKRFESHIECRK 51  
|||||  
  
RESULT 26  
ID Q8BC96 PRELIMINARY; PRT; 72 AA.  
AC Q8BC96;  
DT 01-MAR-2003 (TremBLrel. 23, Created)  
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HAP300;  
RA Villar L.M., Lampe E., Meyer A., Gaspar A.M.C.;  
RT "Genetic variability of hepatitis A virus strains isolated during an  
RT outbreak in a public school in Rio de Janeiro, Brazil.";  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF538727; AAN15169.1; -  
FT NON\_TER 1 72  
FT NON\_TER 72 72  
SQ SEQUENCE 72 AA; 8425 MW; 06C6DAE2309AF3E4 CRC64;  
  
Query Match 100.0%; Score 109; DB 12; Length 72;  
Best Local Similarity 100.0%; Pred. No. 7.5e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VDDPRSEEDKRFESHIECRK 20  
DB 32 VDDPRSEEDKRFESHIECRK 51  
|||||  
  
RESULT 27  
ID Q9QCT0 PRELIMINARY; PRT; 97 AA.  
AC Q9QCT0;  
DT 01-MAY-2000 (TremBLrel. 13, Created)  
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Capsid protein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MIB4022;  
RA Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;  
RT "A Multistate, Foodborne Outbreak of Hepatitis A.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF199046; AAF08937.1; -  
FT NON\_TER 1 97  
FT NON\_TER 97 97  
SQ SEQUENCE 97 AA; 11282 MW; 5ECB8F22812A62B9 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECKR 20  
Db 47 VDDPRSEEDKRFESHIECKR 66

RESULT 28  
Q9QCT1 PRELIMINARY; PRT; 97 AA.  
AC Q9QCT1; (TREMELrel. 13, Created)  
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Capsid protein (Fragment).

OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1] SEQUENCE FROM N.A.  
RP STRAIN=MIB35;  
RC Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;  
RA "A Multistate, Foodborne Outbreak of Hepatitis A";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF199045; AAF08935.1; -;  
DR PIR; PQ0427; PQ0428.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
FT NON\_TER 1  
FT NON\_TER 97  
SQ SEQUENCE 97 AA; 11264 MW; 57C216228E259246 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECKR 20  
Db 47 VDDPRSEEDKRFESHIECKR 66

RESULT 29  
Q9QCT2 PRELIMINARY; PRT; 97 AA.  
AC Q9QCT2; (TREMELrel. 13, Created)  
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Capsid protein (Fragment).

OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1] SEQUENCE FROM N.A.  
RP STRAIN=MIB35;  
RC Nainan O.V., Hutin Y.J., Margolis H.S., Alter M.J.;  
RA "A Multistate, Foodborne Outbreak of Hepatitis A";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF199044; AAF08935.1; -;  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
FT NON\_TER 1  
FT NON\_TER 97  
SQ SEQUENCE 97 AA; 11282 MW; 5ECB8F22812A62B9 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECKR 20  
Db 47 VDDPRSEEDKRFESHIECKR 66

RESULT 30  
Q92941 PRELIMINARY; PRT; 116 AA.  
AC Q92941; (TREMELrel. 08, Created)  
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1] SEQUENCE FROM N.A.  
RP STRAIN=NOR-1;  
RC MEDLINE=97442186; PubMed=9298735;  
RX Grinde B., Stene-Johansen K., Sharma B., Hoel T., Jensenius M.,  
RA Skaug K.;  
RT "Characterisation of an epidemic of hepatitis A virus involving  
RT intravenous drug abusers--infection by needle sharing?";  
RL J. Med. Virol. 53:69-75 (1997).  
DR EMBL; AF050223; AAC05516.1; -;  
DR EMBL; AF050231; AAC05524.1; -;  
DR EMBL; AF050229; AAC05522.1; -;  
DR EMBL; AF050230; AAC05523.1; -;  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
KW Polypeptide.  
FT NON\_TER 1  
FT NON\_TER 116  
SQ SEQUENCE 116 AA; 13479 MW; 4726C6A1E7EBA64 CRC64;

Query Match 100.0%; Score 109; DB 12; Length 116;  
Best Local Similarity 100.0%; Pred. No. 1.2e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECKR 20  
Db 52 VDDPRSEEDKRFESHIECKR 71

Search completed: May 11, 2004, 13:40:21  
Job time : 36 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:34:17 ; Search time 13.5 Seconds  
(without alignments)  
76.483 Million cell updates/sec

Title: US-09-171-432A-38

Perfect score: 109

Sequence: 1 VDDPRSEDKRPFESIECRK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/prodata/2/1aa/6CTUS\_COMB.pep:\*
- 6: /cgn2\_6/prodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	1091	6	Patent No. 5516630
2	109	100.0	2227	3	Sequence 2, Appli
3	109	100.0	2227	3	Sequence 4, Appli
4	109	100.0	2227	3	Sequence 6, Appli
5	109	100.0	2227	3	Sequence 2, Appli
6	109	100.0	2227	3	Sequence 4, Appli
7	109	100.0	2227	3	Sequence 2, Appli
8	109	100.0	2227	4	Sequence 2, Appli
9	109	100.0	2227	4	Sequence 4, Appli
10	109	100.0	2227	4	Sequence 6, Appli
11	109	100.0	2227	4	Sequence 2, Appli
12	109	100.0	2227	4	Sequence 4, Appli
13	109	100.0	2227	4	Sequence 6, Appli
14	109	97.2	2227	4	Sequence 2, Appli
15	81	74.3	839	1	Sequence 12, Appli
16	48.5	44.5	112	4	Sequence 3240, Ap
17	46	42.2	567	4	Sequence 21426, A
18	44.5	40.8	1334	2	Sequence 2, Appli
19	44.5	40.8	1334	3	Sequence 2, Appli
20	44	40.4	185	4	Sequence 30750, A
21	44	40.4	496	4	Sequence 31407, A
22	43.5	39.9	818	4	Sequence 5208, Ap
23	43	39.4	780	2	Sequence 2, Appli
24	42	38.5	180	6	Patent No. 5273901
25	42	38.5	180	6	Patent No. 5482709
26	42	38.5	560	2	Sequence 5, Appli
27	42	38.5	786	4	Sequence 23, Appli

28	42	38.5	786	4	US-09-291-417D-23	Sequence 23, Appli
29	42	38.5	787	4	US-09-688-188B-151	Sequence 151, App
30	42	38.5	787	4	US-09-291-417D-151	Sequence 151, App
31	42	38.5	1001	4	US-09-688-188B-31	Sequence 31, Appl
32	42	38.5	1001	4	US-09-291-417D-31	Sequence 31, Appl
33	41	37.6	168	4	US-09-134-001C-4664	Sequence 4664, A
34	41	37.6	453	4	US-09-252-991A-31268	Sequence 31268, A
35	41	37.6	901	3	US-08-936-135-22	Sequence 22, Appl
36	41	37.6	901	4	US-08-936-135-22	Sequence 22, Appl
37	41	37.6	906	3	US-08-936-135-24	Sequence 24, Appl
38	41	37.6	906	4	US-09-439-711C-24	Sequence 24, Appl
39	41	37.6	909	3	US-08-936-135-8	Sequence 8, Appli
40	41	37.6	909	3	US-08-936-135-10	Sequence 10, Appl
41	41	37.6	909	3	US-08-936-135-18	Sequence 18, Appl
42	41	37.6	909	4	US-09-439-711C-8	Sequence 8, Appli
43	41	37.6	909	4	US-09-439-711C-10	Sequence 10, Appl
44	41	37.6	914	3	US-09-439-711C-18	Sequence 18, Appl
45	41	37.6	914	3	US-08-936-135-12	Sequence 12, Appl
46	41	37.6	914	4	US-09-439-711C-12	Sequence 12, Appl
47	41	37.6	925	4	US-09-116-473-2	Sequence 2, Appli
48	41	37.6	926	3	US-08-936-135-14	Sequence 14, Appl
49	41	37.6	926	3	US-08-936-135-20	Sequence 20, Appl
50	41	37.6	926	4	US-09-439-711C-14	Sequence 14, Appl
51	41	37.6	926	4	US-09-439-711C-20	Sequence 20, Appl
52	41	37.6	931	3	US-08-936-135-16	Sequence 16, Appl
53	41	37.6	931	4	US-09-439-711C-16	Sequence 16, Appl
54	41	37.6	931	4	US-09-583-638-4	Sequence 4, Appli
55	41	37.6	1805	1	US-07-853-913-2	Sequence 2, Appli
56	40.5	37.2	434	4	US-09-489-039A-8188	Sequence 8188, Ap
57	40	36.7	205	4	US-09-252-991A-18606	Sequence 18606, A
58	40	36.7	275	2	US-08-900-565-4	Sequence 4, Appli
59	40	36.7	275	4	US-09-149-534-4	Sequence 4, Appli
60	40	36.7	295	4	US-09-252-991A-20947	Sequence 20947, A
61	40	36.7	304	4	US-09-763-565-2	Sequence 2, Appli
62	40	36.7	324	4	US-09-763-565-4	Sequence 4, Appli
63	40	36.7	324	4	US-09-252-991A-29070	Sequence 29070, A
64	40	36.7	335	2	US-08-379-556A-10	Sequence 10, Appl
65	40	36.7	407	3	US-08-989-370-5	Sequence 5, Appli
66	40	36.7	407	4	US-09-398-169-5	Sequence 5, Appli
67	40	36.7	951	4	US-09-398-169-5	Sequence 5, Appli
68	39	35.8	70	4	US-09-621-976-7468	Sequence 7468, Ap
69	39	35.8	146	3	US-08-792-013-7	Sequence 7, Appli
70	39	35.8	147	1	US-08-171-385-5	Sequence 5, Appli
71	39	35.8	147	3	US-08-361-441B-5	Sequence 5, Appli
72	39	35.8	173	4	US-09-252-991A-21967	Sequence 21967, A
73	39	35.8	181	4	US-09-134-001C-2892	Sequence 2892, Ap
74	39	35.8	244	4	US-09-543-681A-7587	Sequence 7587, Ap
75	39	35.8	298	4	US-09-424-349A-7	Sequence 7, Appli
76	39	35.8	418	4	US-09-252-991A-20665	Sequence 20665, A
77	39	35.8	440	4	US-09-077-955-26	Sequence 26, Appl
78	39	35.8	444	4	US-09-424-283-2	Sequence 2, Appli
79	39	35.8	445	4	US-09-328-352-4631	Sequence 4631, Ap
80	39	35.8	448	4	US-09-252-991A-23910	Sequence 23910, A
81	39	35.8	472	4	US-09-252-991A-17633	Sequence 17633, A
82	39	35.8	504	4	US-09-252-991A-29537	Sequence 29537, A
83	39	35.8	524	4	US-09-424-283-1	Sequence 1, Appli
84	39	35.8	660	4	US-09-252-991A-29885	Sequence 29885, A
85	39	35.8	748	4	US-09-688-188B-24	Sequence 24, Appl
86	39	35.8	748	4	US-09-291-417D-24	Sequence 24, Appl
87	39	35.8	1001	3	US-09-060-410-2	Sequence 2, Appli
88	39	35.8	1001	4	US-09-723-458-2	Sequence 2, Appli
89	39	35.8	1012	4	US-08-811-481-16	Sequence 16, Appl
90	39	35.8	1012	4	US-09-876-527-16	Sequence 16, Appl
91	39	35.8	2165	4	US-09-800-729-155	Sequence 155, App
92	38.5	35.3	373	4	US-09-597-771-10	Sequence 10, Appl
93	38.5	35.3	380	4	US-09-254-776B-80	Sequence 80, Appl
94	38.5	35.3	423	4	US-09-328-352-4348	Sequence 4348, Ap
95	38.5	35.3	433	4	US-09-252-991A-27162	Sequence 27162, A
96	38.5	35.3	544	3	US-08-591-685-9	Sequence 9, Appli
97	38.5	35.3	1349	2	US-08-612-734B-2	Sequence 2, Appli
98	38	34.9	16	4	US-09-485-077A-18	Sequence 18, Appl
99	38	34.9	161	4	US-09-489-039A-13372	Sequence 13372, A
100	38	34.9	262	4	US-09-252-991A-25990	Sequence 25990, A

## ALIGNMENTS

RESULT 1  
5516630-2  
; Patent No. 5516630  
; APPLICANT: TITCHURST, JOHN R.; BALTIMORE, DAVID; FEINSTONE,  
; STEPHEN M.; PURCELL, ROBERT H.; RACANELLO, VINCENT R.;  
; BAROUDY, BAIGIE M.  
; TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION NUMBER: US/07/788,262  
; FILING DATE: 06-NOV-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 256,135  
; FILING DATE: 06-OCT-1988  
; APPLICATION NUMBER: 654,942  
; FILING DATE: 27-SEP-1984  
; APPLICATION NUMBER: 537,911  
; FILING DATE: 30-SEP-1983  
; SEQ ID NO:2:  
; LENGTH: 1091

Query Match 100.0%; Score 109; DB 6; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECRK 20  
Db 1016 VDDPRSEEDKRFESHIECRK 1035

RESULT 2  
US-08-475-886-2  
; Sequence 2, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-08-475-886-2

Query Match 100.0%; Score 109; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECRK 20  
Db 779 VDDPRSEEDKRFESHIECRK 798

RESULT 3  
US-08-475-886-4  
; Sequence 4, Application US/08475886A

; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175  
US-08-475-886-4

Query Match 100.0%; Score 109; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECRK 20  
Db 779 VDDPRSEEDKRFESHIECRK 798

RESULT 4  
US-08-475-886-6  
; Sequence 6, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-08-475-886-6

Query Match 100.0%; Score 109; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECRK 20  
Db 779 VDDPRSEEDKRFESHIECRK 798

RESULT 5  
US-08-397-232-2  
; Sequence 2, Application US/08397232A  
; Patent No. 6180110  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H

APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US1  
; CURRENT APPLICATION NUMBER: US/08/397,232A  
; EARLIER FILING DATE: 1995-04-17  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: PCT/US93/08610  
; EARLIER FILING DATE: 1993-09-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-08-397-232-2

Query Match 100.0%; Score 109; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
Db 779 VDDPRSEEDKRFESHIECKR 798

RESULT 6  
US-08-397-232-4  
; Sequence 4, Application US/08397232A  
; Patent No. 6180110  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US1  
; CURRENT APPLICATION NUMBER: US/08/397,232A  
; CURRENT FILING DATE: 1995-04-17  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: PCT/US93/08610  
; EARLIER FILING DATE: 1993-09-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-08-397-232-4

Query Match 100.0%; Score 109; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
Db 779 VDDPRSEEDKRFESHIECKR 798

RESULT 7  
US-09-171-387-2  
; Sequence 2, Application US/09171387  
; Patent No. 6280734  
; GENERAL INFORMATION:  
; APPLICANT: RAYCHAUDHURI, GOPA;  
; EMERSON, SUZANNE, U.;  
; PURCELL, ROBERT, H.  
; TITLE OF INVENTION: SIMIAN-HUMAN HAV  
; HAVING A CHIMERIC 2C PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.

STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MICROSOFT WORD 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/171,387  
FILING DATE: 24-Mar-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/06506  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US60/015,642  
FILING DATE: 19-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: William S. Feller  
REGISTRATION NUMBER: 26,728  
REFERENCE/DOCKET NUMBER: 2026-4229US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 2  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2227 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-171-387-2

Query Match 100.0%; Score 109; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
Db 779 VDDPRSEEDKRFESHIECKR 798

RESULT 8  
US-09-653-499-2  
; Sequence 2, Application US/09653499  
; Patent No. 6423318  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/09/653,499  
; CURRENT FILING DATE: 2000-08-31  
; PRIOR APPLICATION NUMBER: 08/475,886  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-09-653-499-2

Query Match 100.0%; Score 109; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHIECRK 20  
|||  
db 779 VDDPRSEEDKRFESHIECRK 798

```
Query Match      100.0%; Score 109; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.6e-08;
Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



RESULT 13  
US-10-135-988-6  
; Sequence 6, Application US/10135988  
; Patent No. 6680060  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264362US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-10-135-988-6

Query Match 100.0%; Score 109; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
|||||:|||||:|||||  
Db 779 VDDPRSEEDKRFESHIECKR 798

RESULT 14  
US-10-104-966-12  
; Sequence 12, Application US/10104966  
; Patent No. 6680059  
; GENERAL INFORMATION:  
; APPLICANT: Matti Sallberg  
; APPLICANT: Catharina Hultgren  
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: TRIPEP.23AUSC1  
; CURRENT APPLICATION NUMBER: US/10/104,966  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/705,547  
; PRIOR FILING DATE: 2000-11-03  
; PRIOR APPLICATION NUMBER: 60/229,175  
; PRIOR FILING DATE: 2000-08-29  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Hepatitis A virus sequence  
US-10-104-966-12

Query Match 97.2%; Score 106; DB 4; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 7.5e-08;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIECKR 20  
|||||:|||||:|||||  
Db 779 VDDPRSEEDKRFESHIECKR 798

RESULT 15  
US-08-087-016-2  
; Sequence 2, Application US/08087016  
; Patent No. 5430135

; GENERAL INFORMATION:  
; APPLICANT: NAINAN, OMAYA V.  
; APPLICANT: MARGOLIS, HAROLD S.  
; APPLICANT: ROBERTSON, BETTY H.  
; APPLICANT: BRINTON, MARGO H.  
; APPLICANT: EBERT, JAMES W.  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1615 L Street N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/087,016  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/678,828  
; FILING DATE: 03-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26,581  
; REFERENCE/DOCKET NUMBER: WTS/5683/83834/SRL  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 839 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-087-016-2

Query Match 74.3%; Score 81; DB 1; Length 839;  
Best Local Similarity 82.4%; Pred. No. 0.00019;  
Matches 14; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHIE 17  
|||||:|||||:|||||  
Db 779 VDDPRDEDERFESHIE 795

RESULT 16  
US-09-540-236-3240  
; Sequence 3240, Application US/09540236  
; Patent No. 6673910  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARI  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2005-001  
; CURRENT APPLICATION NUMBER: US/09/540,236  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 3840  
; SEQ ID NO 3240  
; LENGTH: 112  
; TYPE: PRT  
; ORGANISM: M.catarrhalis  
US-09-540-236-3240

Query Match 44.5%; Score 48.5; DB 4; Length 112;  
Best Local Similarity 66.7%; Pred. No. 2.4;  
Matches 10; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 1 VDDPRSEEDKRFESH 15  
||||| : : : : :  
Db 50 VDDP-NNSDERFESH 63

RESULT 17

US-09-252-991A-21426  
; Sequence 21426, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21426  
; LENGTH: 567  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21426

Query Match 42.2%; Score 46; DB 4; Length 567;  
Best Local Similarity 45.0%; Pred. No. 33;  
Matches 9; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

Qy 2 DDPREED--XRFESHIECR 19  
||||| : : : : :  
Db 367 DDPRAVENRRRLTERLECR 386

RESULT 18

US-08-996-545-2  
; Sequence 2, Application US/08996545  
; Patent No. 5928898  
; GENERAL INFORMATION:  
; APPLICANT: Skatrud, Paul L.  
; APPLICANT: de Waard, Maarten A.  
; APPLICANT: Peery, Robert B.  
; APPLICANT: Andrade, Alan C.  
; TITLE OF INVENTION: Multiple Drug Resistance Gene atRD of  
; TITLE OF INVENTION: Aspergillus nidulans  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: U.S.  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/996,545  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/996,545  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Webster, Thomas D.  
; REGISTRATION NUMBER: 39,872  
; REFERENCE/DOCKET NUMBER: X-11766  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3334  
; TELEFAX: 317-276-2763  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1334 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-996-545-2

Query Match 40.8%; Score 44.5; DB 3; Length 1334;  
Best Local Similarity 55.0%; Pred. No. 1.4e+02;  
Matches 11; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

Qy 3 DDPREEDK---RPFESHIECR 19  
||||| : : : : :  
Db 413 DPYSNEGKTLDFEGHIELR 432

RESULT 20

US-09-252-991A-30750

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1334 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-996-545-2

Query Match 40.8%; Score 44.5; DB 2; Length 1334;  
Best Local Similarity 55.0%; Pred. No. 1.4e+02;  
Matches 11; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

Qy 3 DDPREEDK---RPFESHIECR 19  
||||| : : : : :  
Db 413 DPYSNEGKTLDFEGHIELR 432

RESULT 19

US-09-328-320-2  
; Sequence 2, Application US/09328320  
; Patent No. 6228615  
; GENERAL INFORMATION:  
; APPLICANT: Skatrud, Paul L.  
; APPLICANT: de Waard, Maarten A.  
; APPLICANT: Peery, Robert B.  
; APPLICANT: Andrade, Alan C.  
; TITLE OF INVENTION: Multiple Drug Resistance Gene atRD of  
; TITLE OF INVENTION: Aspergillus nidulans  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: U.S.  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/328,320  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/996,545  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Webster, Thomas D.  
; REGISTRATION NUMBER: 39,872  
; REFERENCE/DOCKET NUMBER: X-11766  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3334  
; TELEFAX: 317-276-2763  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1334 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-328-320-2

Query Match 40.8%; Score 44.5; DB 3; Length 1334;  
Best Local Similarity 55.0%; Pred. No. 1.4e+02;  
Matches 11; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

Qy 3 DDPREEDK---RPFESHIECR 19  
||||| : : : : :  
Db 413 DPYSNEGKTLDFEGHIELR 432



```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/559,492
; FILING DATE: 15-NOV-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-873-8794
; TELEFAX: 404-873-8795
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 560 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-559-492-5

Query Match 38.5%; Score 42; DB 2; Length 560;
Best Local Similarity 47.1%; Pred.No.1.4e+02;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 3 DPRSEDKRFESHIECR 19
DB 29 DPELTSQSGASHIDCR 45

RESULT 27
US-09-688-188B-23
; Sequence 23, Application US/09688188B
; Patent No. 6656716
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0328
; CURRENT APPLICATION NUMBER: US/09/688,188B
; CURRENT FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 09/291,417
; PRIOR FILING DATE: 1999-04-14
; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-688-188B-23

Query Match 38.5%; Score 42; DB 4; Length 786;
Best Local Similarity 40.0%; Pred.No. 2e+02;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 6 SEEDKRFESHIECRK 20
DB 315 SNEKKFQFHQAQQ 329

RESULT 28
US-09-291-417D-23
; Sequence 23, Application US/09291417D
; Patent No. 6680170
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY

```

APPLICANT: MARTINEZ, RICARDO  
APPLICANT: WHYTE, DAVID  
TITLE OF INVENTION: ST2D-RELATED PROTEIN KINASES  
FILE REFERENCE: 038602/0329  
CURRENT APPLICATION NUMBER: US/09/291,417D  
CURRENT FILING DATE: 1995-04-13  
PRIOR APPLICATION NUMBER: 60/081,784  
PRIOR FILING DATE: 1998-04-14  
NUMBER OF SEQ IDS NOS: 155

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:37:37 ; Search time 34.25 Seconds  
(without alignments)

162.083 Million cell updates/sec

Title: US-09-171-432A-38

Perfect score: 109

Sequence: 1 VDDPRSEDKRFESHIECKR 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	20	10	US-09-171-432A-38
2	109	100.0	2227	13	US-10-135-988-2
3	109	100.0	2227	13	US-10-135-988-4
4	109	100.0	2227	13	US-10-135-988-6
5	106	97.2	836	14	US-10-272-459-40
6	106	97.2	980	14	US-10-272-459-41
7	106	97.2	2227	9	US-09-929-955-12
8	106	97.2	2227	13	US-10-104-966-12
9	69	63.3	20	10	US-09-171-432A-37
10	66	60.6	302	14	US-10-272-459-44
11	51	45.8	1548	15	US-10-369-493-6347
12	49	45.0	1661	10	US-09-842-758-42
13	49	45.0	1661	12	US-10-174-333-42
14	47	43.1	352	12	US-10-425-114-65499
15	46	42.2	1663	10	US-09-842-758-43
					Sequence 38, Appl
					Sequence 2, Appl
					Sequence 4, Appl
					Sequence 6, Appl
					Sequence 40, Appl
					Sequence 41, Appl
					Sequence 12, Appl
					Sequence 12, Appl
					Sequence 37, Appl
					Sequence 44, Appl
					Sequence 6347, Ap
					Sequence 42, Appl
					Sequence 63499, A
					Sequence 43, Appl

16	46	42.2	1663	12	US-10-174-333-43	Sequence 43, Appl
17	45	41.3	57	12	US-10-424-599-189646	Sequence 189646,
18	45	41.3	119	12	US-10-424-599-272156	Sequence 272156,
19	45	41.3	480	15	US-10-369-493-6516	Sequence 6516, Ap
20	45	41.3	586	15	US-10-369-493-959	Sequence 959, App
21	44.5	40.8	1334	9	US-09-758-828-2	Sequence 2, Appl
22	44	40.4	91	12	US-10-424-599-175273	Sequence 175273,
23	44	40.4	91	12	US-10-424-599-175274	Sequence 175274,
24	43.5	39.9	625	12	US-10-427-224-6	Sequence 6, Appl
25	43	39.4	96	12	US-10-424-599-166177	Sequence 166177,
26	43	39.4	223	15	US-10-369-493-18071	Sequence 18071, A
27	43	39.4	422	15	US-10-369-493-3307	Sequence 3307, Ap
28	43	39.4	535	12	US-10-425-114-55238	Sequence 55238, A
29	43	39.4	607	16	US-10-389-566-1498	Sequence 1498, Ap
30	43	39.4	780	14	US-10-168-066-8	Sequence 8, Appl
31	43	39.4	950	12	US-10-424-599-197989	Sequence 197989,
32	43	39.4	963	15	US-10-369-493-5558	Sequence 5558, Ap
33	43	39.4	19608	15	US-10-084-848A-8	Sequence 8, Appl
34	42	38.5	20	10	US-09-171-432A-36	Sequence 36, Appl
35	42	38.5	54	12	US-10-424-599-173280	Sequence 173280,
36	42	38.5	145	12	US-10-276-774-1926	Sequence 1926, Ap
37	42	38.5	174	12	US-10-424-599-209313	Sequence 209313,
38	42	38.5	175	12	US-10-424-599-153364	Sequence 153364,
39	42	38.5	453	12	US-10-296-115-1430	Sequence 1430, Ap
40	42	38.5	639	16	US-10-389-566-1869	Sequence 1869, Ap
41	42	38.5	704	12	US-10-282-122A-77695	Sequence 77695, A
42	42	38.5	722	14	US-10-161-051-193	Sequence 193, App
43	42	38.5	786	10	US-09-291-417-23	Sequence 23, Appl
44	42	38.5	1001	10	US-09-291-417-31	Sequence 31, Appl
45	42	38.5	1005	14	US-10-177-293-254	Sequence 254, App
46	42	38.5	1005	14	US-10-317-835-18	Sequence 18, Appl
47	42	38.5	1166	15	US-10-104-047-2949	Sequence 2949, Ap
48	42	38.5	2781	12	US-10-263-929-122	Sequence 122, App
49	41.5	38.1	230	15	US-10-369-493-35	Sequence 36, Appl
50	41.5	38.1	744	15	US-10-320-797-3279	Sequence 3279, Ap
51	41	37.6	207	12	US-10-425-114-56558	Sequence 56558, A
52	41	37.6	415	12	US-10-282-122A-78017	Sequence 78017, A
53	41	37.6	491	14	US-10-032-585-7161	Sequence 7161, A
54	41	37.6	492	15	US-10-369-493-22799	Sequence 22799, A
55	41	37.6	614	15	US-10-369-493-5915	Sequence 5915, Ap
56	41	37.6	614	15	US-10-369-493-5916	Sequence 5916, Ap
57	41	37.6	630	15	US-10-259-194A-40	Sequence 40, Appl
58	41	37.6	636	12	US-10-282-122A-50816	Sequence 50816, A
59	41	37.6	926	14	US-10-262-538-4	Sequence 4, Appl
60	41	37.6	926	14	US-10-262-538-8	Sequence 8, Appl
61	41	37.6	926	14	US-10-247-671-171	Sequence 171, App
62	41	37.6	931	13	US-10-104-440-4	Sequence 4, Appl
63	41	37.6	931	14	US-10-104-610-4	Sequence 4, Appl
64	41	37.6	1094	15	US-10-369-493-3622	Sequence 3622, Ap
65	41	37.6	2174	14	US-10-087-887-87	Sequence 87, Appl
66	41	37.6	6642	15	US-10-369-493-5013	Sequence 5013, Ap
67	40.5	37.2	127	12	US-10-424-599-271955	Sequence 271955,
68	40.5	37.2	361	12	US-10-282-122A-45342	Sequence 45342, A
69	40.5	37.2	770	14	US-10-153-668-7	Sequence 7, Appl
70	40.5	37.2	770	14	US-10-153-668-452	Sequence 452, App
71	40.5	37.2	770	15	US-10-094-749-2317	Sequence 2317, Ap
72	40	36.7	20	10	US-09-171-432A-39	Sequence 39, Appl
73	40	36.7	76	12	US-10-424-599-212731	Sequence 212731,
74	40	36.7	97	11	US-09-864-408A-1368	Sequence 1368, Ap
75	40	36.7	101	12	US-10-282-122A-48171	Sequence 48171, A
76	40	36.7	120	14	US-10-029-386-28603	Sequence 28603, A
77	40	36.7	139	11	US-09-864-408A-342	Sequence 342, App
78	40	36.7	179	12	US-10-424-599-160385	Sequence 160385,
79	40	36.7	275	12	US-10-024-933-4	Sequence 4, Appl
80	40	36.7	282	12	US-10-425-114-54392	Sequence 54392, A
81	40	36.7	281	12	US-10-424-599-158843	Sequence 158843,
82	40	36.7	300	12	US-10-424-599-159076	Sequence 159076,
83	40	36.7	315	12	US-10-425-114-45209	Sequence 45209, A
84	40	36.7	352	14	US-10-272-459-45	Sequence 45, Appl
85	40	36.7	365	14	US-10-032-585-7160	Sequence 7160, Ap
86	40	36.7	417	15	US-10-369-493-7000	Sequence 7000, Ap
87	40	36.7	471	12	US-10-425-114-40451	Sequence 40451, A
88	40	36.7	527	14	US-10-156-761-8492	Sequence 8492, Ap

Sequence 24, Appl  
Sequence 12840, A  
Sequence 65790, A  
Sequence 5350, A  
Sequence 5350, Ap  
Sequence 253, App  
Sequence 13, Appl  
Sequence 53173, A  
Sequence 11, Appl  
Sequence 11, Appl  
Sequence 46320, A  
Sequence 240236, A

ALIGNMENTS

RESULT 1  
US-09-171-432A-38  
; Sequence 38, Application US/09171432A  
; Publication No. US20030187184A1  
; GENERAL INFORMATION:  
; APPLICANT: Fields, Howard A.  
; APPLICANT: Khudaykov, Yury E.  
; TITLE OF INVENTION: Antigenically Reactive Regions of the  
; TITLE OF INVENTION: Hepatitis A Virus Polyprotein  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick Stockton LLP  
; STREET: 3424 Peachtree Road, N.E.  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30326  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/09171.432A  
; APPLICATION NUMBER: US/09171.432A  
; FILING DATE: 23-NOV-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Greene, Jamie L.  
; REGISTRATION NUMBER: 32,467  
; REFERENCE/DOCKET NUMBER: 03063-0231US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 949-2400  
; TELEFAX: (404) 949-2499  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..20  
; OTHER INFORMATION: /label= YK-1314  
US-09-171-432A-38

Query Match 100.0%; Score 109; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.1e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIECKR 20  
DB 1 VDDPRSEEDKRFESHIECKR 20

RESULT 2  
US-10-135-988-2  
; Sequence 2, Application US/10135988  
; Publication No. US20020176869A1  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-10-135-988-2

Query Match 100.0%; Score 109; DB 13; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIECKR 20  
DB 779 VDDPRSEEDKRFESHIECKR 798

RESULT 3  
US-10-135-988-4  
; Sequence 4, Application US/10135988  
; Publication No. US20020176869A1  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175  
US-10-135-988-4

Query Match 100.0%; Score 109; DB 13; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VDDPRSEEDKRFESHIECKR 20  
DB 779 VDDPRSEEDKRFESHIECKR 798

RESULT 4  
US-10-135-988-6  
; Sequence 6, Application US/10135988

Query Match	Best Local Similarity	Score	DB	Length	Indels	Gaps
100.0%;	100.0%;	109;	DB 13;	Length 2227;	0;	0;
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	VDDPRSEEDKRFESHIECK 20				
DB	779	VDDPRSEEDKRFESHIECK 798				
LENGTH: 2227						
TYPE: PRT						
ORGANISM: Attenuated (4380) HAV, strain HM-175						
US-10-135-988-6						
Query Match	97.2%;	Score 106;	DB 14;	Length 836;		
Best Local Similarity	95.0%;	Pred. No. 3e-07;				
Matches 19;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	VDDPRSEEDKRFESHIECK 20				
DB	779	VDDPRSEEDKRFESHIECK 798				
LENGTH: 836						
TYPE: PRT						
ORGANISM: Artificial Sequence						
FEATURE:						
OTHER INFORMATION: Description of Artificial Sequence: recombinant protein						
US-10-272-459-40						
Query Match	97.2%;	Score 106;	DB 14;	Length 836;		
Best Local Similarity	95.0%;	Pred. No. 3e-07;				
Matches 19;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	VDDPRSEEDKRFESHIECK 20				
DB	779	VDDPRSEEDKRFESHIECK 798				
LENGTH: 836						
TYPE: PRT						
ORGANISM: Artificial Sequence						
FEATURE:						
OTHER INFORMATION: Description of Artificial Sequence: recombinant protein						
US-10-272-459-40						
Query Match	97.2%;	Score 106;	DB 14;	Length 836;		
Best Local Similarity	95.0%;	Pred. No. 3e-07;				
Matches 19;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	VDDPRSEEDKRFESHIECK 20				
DB	779	VDDPRSEEDKRFESHIECK 798				
LENGTH: 836						
TYPE: PRT						
ORGANISM: Artificial Sequence						
FEATURE:						
OTHER INFORMATION: Description of Artificial Sequence: recombinant protein						
US-10-272-459-40						
Query Match	97.2%;	Score 106;	DB 14;	Length 836;		
Best Local Similarity	95.0%;	Pred. No. 3e-07;				
Matches 19;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	VDDPRSEEDKRFESHIECK 20				
DB	779	VDDPRSEEDKRFESHIECK 798				
LENGTH: 836						
TYPE: PRT						
ORGANISM: Artificial Sequence						
FEATURE:						
OTHER INFORMATION: Description of Artificial Sequence: recombinant protein						
US-10-272-459-40						
Query Match	97.2%;	Score 106;	DB 14;	Length 836;		
Best Local Similarity	95.0%;	Pred. No. 3e-07;				
Matches 19;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	VDDPRSEEDKRFESHIECK 20				
DB	779	VDDPRSEEDKRFESHIECK 798				
LENGTH: 836						
TYPE: PRT						
ORGANISM: Artificial Sequence						
FEATURE:						
OTHER INFORMATION: Description of Artificial Sequence: recombinant protein						
US-10-272-459-40						
Query Match	97.2%;	Score 106;	DB 14;	Length 836;		
Best Local Similarity	95.0%;	Pred. No. 3e-07;				
Matches 19;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	VDDPRSEEDKRFESHIECK 20				
DB	779	VDDPRSEEDKRFESHIECK 798				
LENGTH: 836						
TYPE: PRT						
ORGANISM: Artificial Sequence						
FEATURE:						
OTHER INFORMATION: Description of Artificial Sequence: recombinant protein						
US-10-272-459-40						
Query Match	97.2%;	Score 106;	DB 14;	Length 836;		
Best Local Similarity	95.0%;	Pred. No. 3e-07;				
Matches 19;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	VDDPRSEEDKRFESHIECK 20				
DB	779	VDDPRSEEDKRFESHIECK 798				
LENGTH: 836						
TYPE: PRT						
ORGANISM: Artificial Sequence						



; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Hepatitis A virus sequence  
US-10-104-966-12

Query Match 97.2%; Score 106; DB 13; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 8.5e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VDDPRSEEDKRFESHCK 20  
|||||:|||||  
DB 779 VDDPRSEEDRPFESHCCK 798  
|||||:|||||

RESULT 9  
US-09-171-432A-37  
; Sequence 37, Application US/09171432A  
; Publication No. US20030187184A1  
; GENERAL INFORMATION:  
; APPLICANT: Fields, Howard A.  
; APPLICANT: Khudyakov, Yuri E.  
; TITLE OF INVENTION: Antigenically Reactive Regions of the  
; TITLE OF INVENTION: Hepatitis A Virus Polyprotein  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick Stockton LLP  
; STREET: 3424 Peachtree Road, N.E.  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30326  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/171,432A  
; FILING DATE: 23-NOV-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Greene, Jamie L.  
; REGISTRATION NUMBER: 32,467  
; REFERENCE/DOCKET NUMBER: 03063-0231US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 949-2400  
; TELEFAX: (404) 949-2499  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..20  
; OTHER INFORMATION: /label= YK-1313  
US-09-171-432A-37

Query Match 63.3%; Score 69; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VDDPRSEEDKRFEE 13

Db 8 VDDPRSEEDKRFEE 20  
|||||:|||||

RESULT 10  
US-10-272-459-44  
; Sequence 44, Application US/10272459  
; Publication No. US20030124517A1  
; GENERAL INFORMATION:  
; APPLICANT: PICHAUNTES, Sergio  
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; FILE REFERENCE: PP17955.002 / 2301-17955  
; CURRENT APPLICATION NUMBER: US/10/272,459  
; PRIOR FILING DATE: 2002-10-15  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 44  
; LENGTH: 302  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein  
; OTHER INFORMATION: of 33.3 kDa  
US-10-272-459-44

Query Match 60.6%; Score 66; DB 14; Length 302;  
Best Local Similarity 92.3%; Pred. No. 0.095;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0

QY 1 VDDPRSEEDKRFEE 13  
|||||:|||||  
DB 290 VDDPRSEEDRRFE 302  
|||||:|||||

RESULT 11  
US-10-369-493-6347  
; Sequence 6347, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6347  
; LENGTH: 1548  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-6347

Query Match 46.8%; Score 51; DB 15; Length 1548;  
Best Local Similarity 60.0%; Pred. No. 90;  
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0

QY 2 DDPRESEDKRFESH 16  
::|||:|||||  
DB 479 ENPRSEDRNYESTI 493  
::|||:|||||

RESULT 12  
US-09-842-758-42  
; Sequence 42, Application US/09842758  
; Publication No. US2003083244A1  
; GENERAL INFORMATION:

APPLICANT: Vernet, Corine A. M.  
APPLICANT: Fernandes, Elma R.  
APPLICANT: Gerlach, Valerie  
APPLICANT: Shinkets, Richard A.  
APPLICANT: Malyankar, Uriel M.  
APPLICANT: Boldog, Ferenc L.  
APPLICANT: Zerhusen, Bryan D.  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Majumder, Kumud  
APPLICANT: Tchernev, Velizar T.  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Patturajan, Meera  
APPLICANT: Burgess, Catherine E.  
APPLICANT: Gangolli, Esha A.  
APPLICANT: Smithson, Glenda  
APPLICANT: Rastelli, Luca  
APPLICANT: MacDougall, John R.  
APPLICANT: Taupier, Raymond J.  
APPLICANT: Grosse, William M.  
APPLICANT: Edward, Szekeres S.  
APPLICANT: Alsobrook II, John P.  
APPLICANT: Alsobrook II, John P.  
TITLE OF INVENTION: NO. US2003083244A1el Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-783  
CURRENT APPLICATION NUMBER: US/09/842,758  
CURRENT FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/200,158  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: 60/200,613  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,780  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/201,006  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,007  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,236  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,238  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/201,186  
PRIOR FILING DATE: 2000-05-02  
PRIOR APPLICATION NUMBER: 60/201,474  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/201,508  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 60/220,591  
PRIOR FILING DATE: 2000-07-25  
PRIOR APPLICATION NUMBER: 60/232,678  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 60/263,217  
PRIOR FILING DATE: 2001-01-22  
PRIOR APPLICATION NUMBER: 60/265,160  
PRIOR FILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 113  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 42  
LENGTH: 1661  
TYPE: PRT  
ORGANISM: Sus scrofa  
US-09-842-758-42

Query Match 45.0%; Score 49; DB 10; Length 1661;  
Best Local Similarity 42.1%; Pred. No. 1.9e+02;  
Matches 8; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 2 DDPSEEDKRFESHIECRK 20  
Db 1572 DEVQVGERRFISHIKRE 1590

RESULT 13  
US-10-174-333-42  
Sequence 42, Application US/10174333

Publication No. US20040029220A1  
GENERAL INFORMATION:  
APPLICANT: Vernet, Corine A.M.  
APPLICANT: Fernandes, Elma R.  
APPLICANT: Gerlach, Valerie  
APPLICANT: Malyankar, Uriel M.  
APPLICANT: Boldog, Ferenc L.  
APPLICANT: Zerhusen, Bryan D.  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Majumder, Kumud  
APPLICANT: Tchernev, Velizar T.  
APPLICANT: Padigaru, Muralidhara  
APPLICANT: Patturajan, Meera  
APPLICANT: Burgess, Catherine E.  
APPLICANT: Gangolli, Esha A.  
APPLICANT: Smithson, Glenda  
APPLICANT: Rastelli, Luca  
APPLICANT: MacDougall, John R.  
APPLICANT: Taupier, Raymond J.  
APPLICANT: Grosse, William M.  
APPLICANT: Szekeres, Edward S.  
APPLICANT: Alsobrook, John P.  
APPLICANT: Anderson, David W.  
APPLICANT: Guo, Xiaojia (Sasha)  
APPLICANT: Li, Li  
APPLICANT: Zhong, Mei  
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
FILE REFERENCE: 15966-783 CIP1  
CURRENT APPLICATION NUMBER: US/10/174,333  
CURRENT FILING DATE: 2002-06-18  
PRIOR APPLICATION NUMBER: 60/193,664  
PRIOR FILING DATE: 2000-03-31  
PRIOR APPLICATION NUMBER: 60/194,614  
PRIOR FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: 60/195,063  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,066  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,067  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,068  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,069  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,070  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/195,510  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/219,855  
PRIOR FILING DATE: 2000-07-21  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 186  
SOFTWARE: CuraseqList version 0.1  
SEQ ID NO 42  
LENGTH: 1661  
TYPE: PRT  
ORGANISM: Sus scrofa  
US-10-174-333-42

Query Match 45.0%; Score 49; DB 12; Length 1661;  
Best Local Similarity 42.1%; Pred. No. 1.9e+02;  
Matches 8; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 2 DDPSEEDKRFESHIECRK 20  
Db 1572 DEVQVGERRFISHIKRE 1590

RESULT 14  
US-10-425-114-65499  
Sequence 65499, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:

```

; PRIOR APPLICATION NUMBER: 60/201,186
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: 60/201,474
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 60/201,508
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 60/220,591
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/232,678
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 60/263,217
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/265,160
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 1663
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-842-758-43

Query Match 42.2%; Score 46; DB 10; Length 1663;
Best Local Similarity 38.9%; Pred. No. 5.4e+02;
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 2 DDPSEEDKRFESHIECR 19
   |: : : :|:|:|
Db 1574 DEVOAQQRKFISHIKR 1591

RESULT 16
US-10-174-333-43
; Sequence 43, Application US/10174333
; Publication No. US20040029220A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Gangoli, Sasha A.
; APPLICANT: Smithson, Glennda
; APPLICANT: Fastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Grosse, William M.
; APPLICANT: Szekeres, Edward S.
; APPLICANT: Alsobrook, John P.
; APPLICANT: Anderson, David W.
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Li, Li
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-783 CIP1
; CURRENT APPLICATION NUMBER: US/10/174,333
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: 60/193,664
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/194,614
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: 60/195,063
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,066
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,067

```

```

; PRIOR FILLING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,069
; PRIOR FILLING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,069
; PRIOR FILLING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,070
; PRIOR FILLING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,510
; PRIOR FILLING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/219,855
; PRIOR FILLING DATE: 2000-07-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 43
; LENGTH: 1663
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-174-333-43

```

```

Query Match      42.2%; Score 46; DB 12; Length 1653;
Best Local Similarity 38.3%; Pred. NO. 5.4e+02;
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY      2 DDPRSSEDKRPFESHIECR 19
      |: : : : |||
      |: : : : |||

Db      1574 DEVQAGQQRKF-SHIKCR 1591

```

```

RESULT 17
US-10-424-599-189646
; Sequence 189646, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 189646
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_142266C.1.pep
US-10-424-599-189646

```

```

Query Match      41.3%; Score 45; DB 12; Length 57;
Best Local Similarity 69.2%; Pred. NO. 22;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      6 SEEDKRFESHIEC 18
      | | | | | | | |
Db      10 SRVQKRFSLSHIEC 22

```

RESULT 18  
US-10-424-599-272156  
; Sequence 272156, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B



```

; NAME/KEY: unsure
; LOCATION: (1)..(223)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-18071

Query Match 39.4%; Score 43; DB 15; Length 223;
Best Local Similarity 36.8%; Pred. No. 1.8e+02;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 2 DDPRSEEDKRFESHIECKR 20
Db 11 EEPAGDRKRAEAHTETHR 29

RESULT 27
US-10-369-493-3307
; Sequence 3307, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3307
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-3307

Query Match 39.4%; Score 43; DB 15; Length 422;
Best Local Similarity 47.6%; Pred. No. 3.6e+02;
Matches 10; Conservative 3; Mismatches 4; Indels 4; Gaps 1;

Qy 2 DDPRSEEDKRFESHIEC 18
Db 217 DEEAERGFPFGSFESQIEC 237

RESULT 28
US-10-425-114-55238
; Sequence 55238, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 55238
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73019H10_FLI pep
US-10-425-114-55238

Query Match 39.4%; Score 43; DB 12; Length 535;

; PRIOR FILING DATE: 2002-05-01
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-427-224-6

Query Match 39.9%; Score 43.5; DB 12; Length 625;
Best Local Similarity 45.5%; Pred. No. 4.6e+02;
Matches 10; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

Qy 1 VDDPRSEEDKRFES---HIECR 19
Db 214 LDDIPEEDPQYSEAEHVEAR 235

RESULT 25
US-10-424-599-166177
; Sequence 166177, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 166177
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_121073C.1.pep
US-10-424-599-166177

Query Match 39.4%; Score 43; DB 12; Length 96;
Best Local Similarity 50.0%; Pred. No. 75;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VDDPRSEEDKRFESHI 16
Db 19 VDNPSAVDDLAFHSHI 34

RESULT 26
US-10-369-493-18071
; Sequence 18071, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18071
; LENGTH: 223
; TYPE: PRT
; ORGANISM: SPHINGOMONAS
; FEATURE:

```

Best Local Similarity 70.0%; Pred. No. 4.6e+02;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 DDPRSEEDKR 11  
:||||:|  
Db 496 DNPRSQDKR 505

Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 VDDPRSEEDKRFESHIECR 19  
:||||:|:|:|:|:|:|:|  
Db 74 LDDPASQSTIERGKSYLR 92

Search completed: May 11, 2004, 13:44:48  
Job time : 35.25 secs

RESULT 29  
US-10-389-566-1498  
; Sequence 1498, Application US/10389566  
; Publication No. US20040025202A1  
; GENERAL INFORMATION:  
; APPLICANT: Monsanto Technology, LLC  
; APPLICANT: Laurie, Cathy C  
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants  
; FILE REFERENCE: 38-77(52900)D  
; CURRENT APPLICATION NUMBER: US/10/389,566  
; CURRENT FILING DATE: 2003-03-31  
; PRIOR APPLICATION NUMBER: US 60/365,301  
; PRIOR FILING DATE: 2002-03-15  
; PRIOR APPLICATION NUMBER: US 60/391,786  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: US 60/392,018  
; PRIOR FILING DATE: 2002-06-26  
; NUMBER OF SEQ ID NOS: 2459  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 1498  
; LENGTH: 607  
; TYPE: PRT  
; ORGANISM: Fusobacterium nucleatum  
US-10-389-566-1498

Query Match 39.4%; Score 43; DB 16; Length 607;  
Best Local Similarity 41.2%; Pred. No. 5.2e+02;  
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 3 DPRSEEDKRFESHIECR 19  
:||||:|:|  
Db 491 EAAEEDKKFQELVEAR 507

RESULT 30  
US-10-168-066-8  
; Sequence 8, Application US/10168066  
; Publication No. US20030087268A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: YUE, Henry  
; APPLICANT: BANDMAN, Olga  
; APPLICANT: TANG, Y. Tom  
; APPLICANT: HILLMAN, Jennifer L.  
; APPLICANT: LJ, Dyung Aina M.  
; APPLICANT: BAUGHN, Mariah R.  
; TITLE OF INVENTION: HUMAN LYASES AND ASSOCIATED PROTEINS  
; FILE REFERENCE: PF 0759 PCT  
; CURRENT APPLICATION NUMBER: US/10/168,066  
; CURRENT FILING DATE: 2000-12-13  
; PRIOR APPLICATION NUMBER: 60/172,307  
; PRIOR FILING DATE: 1999-12-16  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PERL Program  
; SEQ ID NO 8  
; LENGTH: 780  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030087268A1 3130234  
US-10-168-066-8

Query Match 39.4%; Score 43; DB 14; Length 780;  
Best Local Similarity 42.1%; Pred. No. 6.8e+02;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2004, 13:50:08 ; Search time 10 seconds  
(without alignments)  
192.383 Million cell updates/sec

Title: US-09-171-432A-39

Perfect score: 107

Sequence: 1 SHIECKPYKELRLEVQKOR 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	56	PQ0431	genome polyprotein
2	107	100.0	341	S04137	genome polyprotein
3	107	100.0	852	GNHYHA	genome polyprotein
4	107	100.0	1358	A03905	genome polyprotein
5	107	100.0	2227	GNHYHM	genome polyprotein
6	107	100.0	2227	GNHYHR	genome polyprotein
7	107	100.0	2227	GNHYMK	genome polyprotein
8	107	100.0	2227	GNHYHB	genome polyprotein
9	105	98.1	56	PQ0427	genome polyprotein
10	104	97.2	56	PQ0429	genome polyprotein
11	102	95.3	319	JH0135	genome polyprotein
12	100	93.5	56	PQ0430	genome polyprotein
13	97	90.7	56	PQ0432	genome polyprotein
14	96	89.7	56	PQ0428	genome polyprotein
15	95	88.8	56	PQ0434	genome polyprotein
16	87	81.3	55	PQ0433	genome polyprotein
17	86	80.4	56	PQ0436	genome polyprotein
18	86	80.4	2230	GNHYSA	genome polyprotein
19	84.5	79.0	55	PQ0435	genome polyprotein
20	84.5	79.0	839	GNYS2	genome polyprotein
21	81	75.7	56	PQ0437	genome polyprotein
22	46	43.0	420	S62541	hypothetical prote
23	44	41.1	374	C84C40	hypothetical prote
24	43	40.7	1663	C35T	complement C3 prec
25	43	40.2	332	P95568	probable lipase, 2
26	43	40.2	339	T18926	hypothetical prote
27	43	40.2	343	AH1823	30S ribosomal prot
28	43	40.2	346	T27896	hypothetical prote
29	43	40.2	391	VHNZ3	nucleocapsid prote

30 43 40.2 554 2 T49917  
31 42.5 39.7 140 2 E64364  
32 42 39.3 121 2 T17708  
33 42 39.3 214 2 E82679  
34 42 39.3 286 2 S07532  
35 42 39.3 329 2 T25067  
36 42 39.3 494 2 S62902  
37 42 39.3 723 2 D71091  
38 42 39.3 861 2 B84963  
39 42 39.3 1087 2 T16876  
40 42 39.3 1157 2 T43259  
41 42 39.3 1270 2 T22615  
42 41.5 38.8 286 2 H87506  
43 41.5 38.8 606 2 A12649  
44 41.5 38.8 606 2 H97431  
45 41 38.3 144 2 H96511

hypothetical prote  
hypothetical prote  
hypothetical prote  
chaperone XF1452  
puff 17/9-1 protei  
hypothetical prote  
legumin 2 precursor  
methionine-tRNA li  
DNA topoisomerase  
hypothetical prote  
pyruvate (flavodox  
hypothetical prote  
phosphogluconate d  
phosphogluconate d  
AP2 domain contain

#### ALIGNMENTS

##### RESULT 1

PQ0431  
genome polyprotein - human hepatitis A virus (strain No. 4) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0431  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Wide  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geog  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0431  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>

A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 100.0%; Score 107; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 4e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPYKELRLEVQKOR 20  
|||  
DB 29 SHIECKPYKELRLEVQKOR 48  
|||

##### RESULT 2

S04137  
genome polyprotein - human hepatitis A virus (strain LCDC-1) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
C:Accession: S04137  
R:Andonov, A.P.; Lau, P.C.K.; Chaudhary, R.  
Nucleic Acids Res. 17, 3594, 1989  
A:Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A virus  
A:Reference number: S04137; MUID:89263805; PMID:2542903  
A:Accession: S04137  
A:Molecule type: mRNA  
A:Residues: 1-341 <AND>

A:Cross-references: EMBL:X14666; NID:G62301; PIDN:CAA32794.1; PID:G4377576

C:Genetics:

A:Gene: VP1

C:Superfamily: hepatitis A virus genome polyprotein

C:Keywords: coat protein; polyprotein

F:2-340/Product: coat protein ID (VP1) #status predicted <MAT>

Query Match 100.0%; Score 107; DB 2; Length 341;

Best Local Similarity 100.0%; Pred. No. 2.5e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPYKELRLEVQKOR 20  
|||



Db 296 SHIECRPKYKELRLEVGVKQR 315

## RESULT 3

GNVYHA  
genome polyprotein - human hepatitis A virus (strain CR326) (fragment)  
N/Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A  
C/Species: human hepatitis A virus  
A/Note: host Homo sapiens (man)  
C/Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C/Accession: A03904  
R/Lineage: D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitra, S.W.  
J. Virol. 54, 247-255, 1985  
A/Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.  
A/Reference number: A03904; MUID:85185648; PMID:2985793  
A/Accession: A03904  
A/Molecule type: Genomic RNA  
A/Residues: 1-852 <LIN>  
A/Cross-references: EMBL:MI0033; NID:g329592; PID:AAA45470.1; PID:g329593  
C/Superfamily: hepatitis A virus genome polyprotein  
C/Keywords: coat protein; core protein; polyprotein  
F/1-245/Product: coat protein 1A #status predicted <CIA>  
F/246-491/Product: coat protein 1B #status predicted <CIB>  
F/492-836/Product: coat protein 1C #status predicted <CIC>  
F/837-852/Product: core protein 2A (fragment) #status predicted <C2A>

Query Match 100.0%; Score 107; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 6.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGVKQR 20

Db 792 SHIECRPKYKELRLEVGVKQR 811

## RESULT 4

A03905  
genome polyprotein (version 2) - human hepatitis A virus (fragments)  
N/Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; protein 3  
C/Species: human hepatitis A virus  
C/Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 15-Nov-1996  
C/Accession: A03905  
R/Baroudy, B.M.; Ticehurst, J.R.; Mile, T.A.; Maizel Jr., J.V.; Purcell, R.H.; Feinstone  
Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985  
A/Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA  
A/Reference number: A03905; MUID:85166289; PMID:2984684  
A/Accession: A03905  
A/Molecule type: Genomic RNA  
A/Residues: 1-1358 <BAR>  
C/Superfamily: hepatitis A virus genome polyprotein  
C/Keywords: coat protein; core protein; polyprotein  
F/1-245/Product: coat protein 1A #status predicted <CIA>  
F/246-491/Product: coat protein 1B #status predicted <CIB>  
F/492-836/Product: coat protein 1C #status predicted <CIC>  
F/837-854/Product: core protein 2A (fragment) #status predicted <C2A>  
F/855-1358/Product: protein 3D (RNA polymerase) (fragment) #status predicted <C3D>

Query Match 100.0%; Score 107; DB 2; Length 1358;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGVKQR 20

Db 792 SHIECRPKYKELRLEVGVKQR 811

## RESULT 5

GNVYHM  
genome polyprotein - human hepatitis A virus (strain HM-175, wild type)  
N/Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core protein 3  
B; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C/Species: human hepatitis A virus  
A/Note: host Homo sapiens (man)

C/Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999  
C/Accession: A25981  
R/Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.  
J. Virol. 61, 50-59, 1987  
A/Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with c  
A/Reference number: A25981; MUID:87061253; PMID:3023706  
A/Accession: A25981  
A/Molecule type: Genomic RNA  
A/Residues: 1-2227 <COH>  
A/Cross-references: EMBL:MI4707; NID:g329582; PID:AAA45465.1; PID:g329583  
C/Superfamily: hepatitis A virus genome polyprotein  
C/Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltran  
F/1-23/Product: coat protein 1A #status predicted <VP4>  
F/24-245/Product: coat protein 1B #status predicted <VP2>  
F/246-491/Product: coat protein 1C #status predicted <VP3>  
F/492-791/Product: coat protein 1D #status predicted <VP1>  
F/792-980/Product: core protein 2A #status predicted <C2A>  
F/981-1087/Product: core protein 2B #status predicted <C2B>  
F/1088-1422/Product: core protein 2C #status predicted <C2C>  
F/1423-1496/Product: protein 3A #status predicted <C3A>  
F/1497-1519/Product: protein 3B #status predicted <C3B>  
F/1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F/1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 107; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGVKQR 20

Db 792 SHIECRPKYKELRLEVGVKQR 811

## RESULT 6

GNVYHR  
genome polyprotein - human hepatitis A virus  
N/Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core pr  
NA polymerase (EC 2.7.7.48), protein 3D  
C/Species: human hepatitis A virus  
A/Note: host Homo sapiens (man)  
C/Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C/Accession: A03903  
R/Najarian, R.J.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van Nes  
Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985  
A/Title: Primary structure and gene organization of human hepatitis A virus.  
A/Reference number: A03903; MUID:85190549; PMID:2986127  
A/Accession: A03903  
A/Molecule type: Genomic RNA  
A/Residues: 1-2227 <NAJ>  
A/Cross-references: GB:K02990; NID:g329596; PID:AAA45472.1; PID:g329597  
C/Superfamily: hepatitis A virus genome polyprotein  
C/Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltran  
F/1-245/Product: coat protein 1A #status predicted <CIA>  
F/246-491/Product: coat protein 1B #status predicted <CIB>  
F/492-836/Product: coat protein 1C #status predicted <CIC>  
F/837-980/Product: core protein 2A #status predicted <C2A>  
F/981-1076/Product: core protein 2B #status predicted <C2B>  
F/1077-1422/Product: core protein 2C #status predicted <C2C>  
F/1423-1484/Product: protein 3A #status predicted <C3A>  
F/1485-1507/Product: protein 3B #status predicted <C3B>  
F/1508-1678/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F/1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 107; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGVKQR 20

Db 792 SHIECRPKYKELRLEVGVKQR 811

## RESULT 7

GNMYK  
genome polyprotein - human hepatitis A virus (strain HM-175/7MK-5, attenuated HAV)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core protein 2B; core protein 2C; core protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Jul-1999  
C:Accession: A94149; A25914; A94508  
R:Cohen, J.L.; Rosenblum, B.; Ticehurst, J.R.; Daemer, R.J.; Feinstein, S.M.; Purcell, R.  
Proc. Natl. Acad. Sci. U.S.A. 84, 2497-2501, 1987  
A:Title: Complete nucleotide sequence of an attenuated hepatitis A virus: comparison with reference sequence  
A:Reference number: A94149; MUID:87175701; PMID:3031686  
A:Accession: A94149  
A:Status: nucleic acid sequence not shown  
A:Molecule type: Genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M16632; NID:G329594; PIDN:AAA45471.1; PID:G329595  
A:Note: submitted to GenBank, August 1987  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltransferase  
F:1-245/Product: coat protein 1A #status predicted <P1A>  
F:246-491/Product: coat protein 1B #status predicted <P1B>  
F:492-836/Product: coat protein 1C #status predicted <P1C>  
F:837-980/Product: coat protein 2A #status predicted <P2A>  
F:981-1076/Product: coat protein 2B #status predicted <P2B>  
F:1077-1422/Product: core protein 2C #status predicted <P2C>  
F:1423-1484/Product: protein 3A #status predicted <P3A>  
F:1485-1507/Product: protein 3B #status predicted <P3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>  
Query Match 100.0%; Score 107; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPKYKELRLEVGKQR 20  
|||||  
Db 792 SHIECRKPKYKELRLEVGKQR 811

RESULT 8  
GNMYB  
genome polyprotein - human hepatitis A virus (strain MBB)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core protein 2A; core protein 2B; core protein 2C; core protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 10-May-1996  
C:Accession: J50303  
R:Paul, A.V.; Tada, H.; von der Helm, K.; Wissel, T.; Kiehn, R.; Wimmer, E.; Deinhardt, V.  
Virus Res. 8, 153-171, 1987  
A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (isolated from a patient with acute hepatitis A)  
A:Reference number: J50303; MUID:88045071; PMID:2823500  
A:Accession: J50303  
A:Molecule type: Genomic RNA  
A:Residues: 1-2227 <PAU>  
A:Cross-references: EMBL:M20273  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; genome-linked protein; hydrophobicity  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-246/Product: coat protein 1B #status predicted <VP2>  
F:247-491/Product: coat protein 1C #status predicted <VP3>  
F:492-836/Product: coat protein 1D #status predicted <VP1>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1108/Product: core protein 2B #status predicted <P2B>  
F:1109-1438/Product: core protein 2C #status predicted <P2C>  
F:1439-1496/Product: core protein 3A #status predicted <P3A>  
F:1497-1519/Product: genome-linked protein VPg #status predicted <VPg>  
F:1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>  
Query Match 100.0%; Score 107; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPKYKELRLEVGKQR 20  
|||||  
Db 792 SHIECRKPKYKELRLEVGKQR 811

RESULT 9  
PQ0427  
genome polyprotein - human hepatitis A virus (strain EP-35.730) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0427  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Wide  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geographic regions  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0427  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
C:Comment: This protein is from the VP1/2A Junction region.  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
Query Match 98.1%; Score 105; DB 2; Length 56;  
Best Local Similarity 95.0%; Pred. No. 8.3e-10;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPKYKELRLEVGKQR 20  
|||||  
Db 29 SHIECRKPKYKELRLEVGKQR 48

RESULT 10  
PQ0429  
genome polyprotein - human hepatitis A virus (strain PRC16) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0429  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Wide  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geographic regions  
A:Reference number: PQ0429; MUID:92300330; PMID:1318940  
A:Accession: PQ0429  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
Query Match 97.2%; Score 104; DB 2; Length 56;  
Best Local Similarity 95.0%; Pred. No. 1.2e-09;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPKYKELRLEVGKQR 20  
|||||  
Db 29 SHIECRKPKYKELRLEVGKQR 48

RESULT 11  
JH0135  
genome polyprotein - human hepatitis A virus (strain MS-1) (fragment)  
N:Contains: amino end of core protein 2A; coat protein 1D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 08-Apr-1994  
C:Accession: JH0135  
R:Robertson, B.H.; Brown, V.K.; Bradley, D.W.  
Virus Res. 8, 303-316, 1987  
A:Title: Nucleic acid sequence of the VP1 region of attenuated MS-1 hepatitis A virus.  
A:Reference number: JH0135; MUID:88129044; PMID:2829458  
A:Accession: JH0135  
A:Molecule type: Genomic RNA

A:Residues: 1-319 <ROB>  
A:Cross-references: GB:M22821  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; glycoprotein; polypeptide  
F:1-300/Product: coat protein 1D #status predicted <CPD>  
F:301-319/Product: core protein 2A (fragment) #status predicted <C2A>  
F:237/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 95.3%; Score 102; DB 2; Length 319;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGKQ 19  
|||||  
Db 301 SHIECRKPYKELRLEVGKQ 319  
|||||

RESULT 12  
PQ0430  
genome polypeptide - human hepatitis A virus (strain S23-1) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0430  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widell  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0430  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; polypeptide

Query Match 93.5%; Score 100; DB 2; Length 56;  
Best Local Similarity 95.0%; Pred. No. 5.2e-09;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGKQ 20  
|||||  
Db 29 SHIECRKPYKELRLEVGKQ 48  
|||||

RESULT 13  
PQ0432  
genome polypeptide - human hepatitis A virus (strain CP-53) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0432  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widell

C:Accession: PQ0432  
R:Robertson, B.H.; Jansen, R.W.; Klanna, B.; Totsuka, A.; Nainan, O.V.; Sisq, G.; Widel  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 0-May-1999

34 41 38.3 492 1 ANKH MOUSE  
35 41 38.3 492 1 ANKH RAT  
36 41 38.3 492 1 ANKH XENIA  
37 41 38.3 840 1 NLG1\_HUMAN  
38 41 38.3 843 1 NLG1\_MOUSE  
39 41 38.3 843 1 NLG1\_MOUSE  
40 41 38.3 955 1 T150\_HUMAN  
41 41 38.3 4543 1 LRP1\_CHICK  
42 41 38.3 4544 1 LRP1\_HUMAN  
43 40.5 37.9 141 1 V16K\_TRVTC  
44 40.5 37.9 274 1 NAGB\_FUSNN  
45 40.5 37.9 1120 1 RPOM\_SCHPO

Q9jhz2 mus musculus  
P58366 rattus norv  
P58367 xenopus lae  
Q8n2q7 homo sapien  
Q99k10 mus musculus  
Q62765 rattus norv  
Q9y2w1 homo sapien  
P98157 gallus gall  
Q07554 homo sapien  
P05076 tobacco rat  
Q8reg1 fusobacteri  
O13993 schizosacch

ALIGNMENTS

RESULT 1  
POLG\_HPAV1  
ID POLG\_HPAV1 STANDARD; PRT; 341 AA.  
AC FI3672; 1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-JAN-1990 (Rel. 13, Last annotation update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polypeptide [Contains: Coat protein VP3 (1C); Coat protein VP1 (1D); Core protein P2A] (Fragment).  
OS Hepatitis A virus (strain LCD-1).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12093;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=89263805; PubMed=2542903;  
RX Andonov A.P., Lau P., Chaudhary R.;  
RT "Nucleotide sequence of the VP1 gene from a Chinese strain of  
hepatitis A virus (HAV)";  
RL Nucleic Acids Res. 17:3594-3594(1989).  
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
each of which is composed of one copy each of proteins VP1, VP2,  
VP3, and VP4.  
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X14666; CAA32794.1; -  
DR PIR; S04137; S04137.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
KW Polypeptide; Coat protein; Core protein.  
FT NON\_TER 1 1  
FT CHAIN <1 1 COAT PROTEIN VP3.  
FT CHAIN 2 340 COAT PROTEIN VP1.  
FT CHAIN 341 >341 CORE PROTEIN P2A.  
FT NON\_TER 341 341  
SQ SEQUENCE 341 AA; 38003 MW; 066918289BF126D5 CRC64;  
Query Match 100.0%; Score 107; DB 1; Length 341;  
Best Local Similarity 100.0%; Pred. No. 6.6e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SHIECRKPYKELRLEVQKQR 20  
Db 296 SHIECRKPYKELRLEVQKQR 315  
RESULT 2  
POLG\_HPAV1  
ID POLG\_HPAV1 STANDARD; PRT; 852 AA.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: March 15, 2004, 13:26:12 ; Search time 5.64706 Seconds  
(without alignments)  
184.415 Million cell updates/sec

Title: US-09-171-432A-39  
Perfect score: 107  
Sequence: 1 SHIECRKPYKELRLEVQKQR 20  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	107	100.0	341	1	POLG_HPAV1
2	107	100.0	852	1	POLG_HPAVC
3	107	100.0	2226	1	POLG_HPAV2
4	107	100.0	2226	1	POLG_HPAV4
5	107	100.0	2226	1	POLG_HPAV8
6	107	100.0	2227	1	POLG_HPAVH
7	107	100.0	2227	1	POLG_HPAVL
8	107	100.0	2227	1	POLG_HPAVM
9	95	88.8	808	1	POLG_HPAVG
10	86	80.4	2230	1	POLG_HPAVS
11	84.5	79.0	839	1	POLG_HPAVT
12	46	43.0	420	1	YAGA_SCHPO
13	45	42.1	1819	1	GCPE_HUMAN
14	43.5	40.7	1663	1	CO3 RAT
15	43	40.2	391	1	NCAP_HRSV1
16	42.5	39.7	140	1	Y517_VETJA
17	42	39.3	210	1	LOLA_XLEFA
18	42	39.3	210	1	LOLA_XLEFT
19	42	39.3	286	1	FU91_SCICO
20	42	39.3	298	1	END4_BACAA
21	42	39.3	298	1	END4_BACCR
22	42	39.3	723	1	SYM_PFRHO
23	42	39.3	861	1	TOPI_BUCAL
24	41	38.3	273	1	RL2_BUCAP
25	41	38.3	298	1	END4_BACHD
26	41	38.3	319	1	URBD_SYNVP
27	41	38.3	364	1	SUCC_VETJA
28	41	38.3	370	1	T2M2_VETJA
29	41	38.3	391	1	NCAP_BRSV3
30	41	38.3	391	1	NCAP_BRSVA
31	41	38.3	391	1	NCAP_BRSVR
32	41	38.3	391	1	NCAP_HRSVA
33	41	38.3	492	1	ANKH_HUMAN

AC P06442; Q83741; Q83742;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein  
DE P2A] (Fragment).  
OS Hepatitis A virus (strain CR326).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85195648; PubMed=2965793;  
RA Lineneyer D.L., Menke J.G., Martin-Gallardo A., Hughes J.V.,  
RA Young A., Mitra S.W.,  
RT "Molecular cloning and partial sequencing of hepatitis A viral cDNA."  
RL J. Virol. 54:247-255(1985).  
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M10033; AAA45470.1; --  
DR PIR; A03904; GNNYHA.  
DR InterPro; IPR008975; Viral cap coat.  
KW Polyprotein; Coat protein; Core protein.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 836 COAT PROTEIN VP1 (PID).  
FT CHAIN 837 >852 CORE PROTEIN P2A.  
FT NON TER 852 852  
SQ SEQUENCE 852 AA; 95563 MW; 73D3ED0AD532820E CRC64;  
Query Match 100.0%; Score 107; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 1.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20  
Db 792 SHIECRPKYKELRLEVGKQR 811

RESULT 3  
POLG HPAV2 STANDARD; PRT; 2226 AA.  
AC P26580;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 24a).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12094;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
RA Cromean T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic

RT recombination.";  
RL J. Virol. 65:2056-2065(1991).  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC (RNA)(N).  
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M59810; AAA45468.1; --  
DR MEROPS; C03.005; --  
DR InterPro; IPR004004; Calici.pol.hel.  
DR InterPro; IPR003003; Cys Ser trypsin.  
DR InterPro; IPR006005; RNA helicase.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR007094; RNA\_pol\_P3vir.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR PRINTS; PR00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICIVIRUSN.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).  
FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;

Query Match 100.0%; Score 107; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 4.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20  
Db 792 SHIECRPKYKELRLEVGKQR 811

RESULT 4  
POLG HPAV4 STANDARD; PRT; 2226 AA.  
AC P26581;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 43C).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12095;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
RA Cromean T., Jansen R.W.;

RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination.";  
RL J. Virol. 65:2056-2065(1991).  
CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC [RNA](N).  
CC -|- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -|- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M59809; AAA45469.1; --  
CC MEROPS; C03.005; --  
CC InterPro; IPR004004; Calici\_pol\_hel.  
CC InterPro; IPR009003; Cys\_Ser\_trypsin.  
CC InterPro; IPR006005; RNA\_helicase.  
CC InterPro; IPR007095; RNA\_pol\_DS\_PS.  
CC InterPro; IPR001205; RNA\_pol\_P3D.  
CC InterPro; IPR007094; RNA\_pol\_PSVir.  
CC InterPro; IPR008975; Viral\_cap\_coat.  
CC Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
CC Pfam; PF00910; RNA\_helicase; 1.  
CC PRINTS; PR00918; CALICIVIRUSNS.  
CC Polyprotein; Coat protein; Core protein; Transferase;  
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1495  
FT CHAIN 1496 1518  
FT CHAIN 1519 1737  
FT CHAIN 1738 2226  
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B9BF75 CRC64;  
Query Match 100.0%; Score 107; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 4.4e-09; Mismatches 0; Indels 0; Gaps 0;  
Matches 20; Conservative 0;  
Qy 1 SHIECRKPKYKELRLEVGKQR 20  
Db 792 SHIECRKPKYKELRLEVGKQR 811  
RESULT 5  
ID POLG\_HPAV8 STANDARD; PRT; 2226 AA.  
AC P26582;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 18f).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12096;  
RN [1]\_TaxID=12098;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;

RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
RA Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination.";  
RL J. Virol. 65:2056-2065(1991).  
CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC [RNA](N).  
CC -|- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -|- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M59808; AAA45467.1; --  
CC PDB; 1QA7; 15-MAY-00.  
CC MEROPS; C03.005; --  
CC InterPro; IPR004004; Calici\_pol\_hel.  
CC InterPro; IPR009003; Cys\_Ser\_trypsin.  
CC InterPro; IPR006005; RNA\_helicase.  
CC InterPro; IPR007095; RNA\_pol\_DS\_PS.  
CC InterPro; IPR001205; RNA\_pol\_P3D.  
CC InterPro; IPR007094; RNA\_pol\_PSVir.  
CC InterPro; IPR008975; Viral\_cap\_coat.  
CC Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
CC Pfam; PF00910; RNA\_helicase; 1.  
CC PRINTS; PR00918; CALICIVIRUSNS.  
CC Polyprotein; Coat protein; Core protein; Transferase;  
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 794  
FT CHAIN 795 900  
FT CHAIN 901 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1495  
FT CHAIN 1496 1518  
FT CHAIN 1519 1737  
FT CHAIN 1738 2226  
SQ SEQUENCE 2226 AA; 251292 MW; 24964A63396C8D6B CRC64;  
Query Match 100.0%; Score 107; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 4.4e-09; Mismatches 0; Indels 0; Gaps 0;  
Matches 20; Conservative 0;  
Qy 1 SHIECRKPKYKELRLEVGKQR 20  
Db 792 SHIECRKPKYKELRLEVGKQR 811  
RESULT 6  
ID POLG\_HPAVH STANDARD; PRT; 2227 AA.  
AC P08617; P06443; Q81082;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain HM-175).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12098;

[1] SEQUENCE FROM N.A.  
RN STRAIN=Wild type;  
RRP MEDLINE=87061253; PubMed=3023706;  
RC Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A.,  
RRA Baroudy B.M.;  
RRC "Complete nucleotide sequence of wild-type hepatitis A virus:  
RRT comparison with different strains of hepatitis A virus and other  
RRT picornaviruses.";  
RRL J. Virol. 61:50-59 (1987).  
[2] SEQUENCE FROM N.A.  
RN STRAIN=Attenuated;  
RRP MEDLINE=87175701; PubMed=3031686;  
RRC Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstein S.M.,  
RRA Purcell R.H.;  
RRC "Complete nucleotide sequence of an attenuated hepatitis A virus:  
RRT comparison with wild-type virus.";  
RRL Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501 (1987).  
[3] SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.  
RN MEDLINE=85166289; PubMed=2984694;  
RRP Baroudy B.M., Ticehurst J.R., Miele T.A., Maizel J.V. Jr.,  
RRA Purcell R.H., Feinstein S.M.;  
RRC "Sequence analysis of hepatitis A virus cDNA coding for capsid  
RRT proteins and RNA polymerase.";  
RRL Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147 (1985).  
- !- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC (RNA) (N).  
CC  
- !- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC  
- !- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC  
- !- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED  
CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.  
CC  
- !- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC  
- !- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT  
CC SHOWN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M14114; AAA45475.1; -;  
DR EMBL; M14707; AAA45465.1; -;  
DR EMBL; M14707; AAA45466.1; ALT\_INIT.  
DR EMBL; M16632; AAA45471.1; -;  
DR FIR; A03905; A03905.  
DR FIR; A25981; GNNYHM.  
DR FIR; A94149; GNNYMK.  
DR FDB; 1HAV; 23-DEC-96.  
DR MEROPS; C03.005; -;  
DR InterPro; IPR004004; Calici\_pol\_hel.  
DR InterPro; IPR009003; Cys\_ser\_tpsin.  
DR InterPro; IPR000605; RNA\_helicase.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR007094; RNA\_pol\_Psvir.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol\_1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICIVIRUSN.  
DR PolyProtein; Coat protein; Core protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.  
RN CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN CORE PROTEIN VP1  
FT CHAIN CORE PROTEIN VP2  
FT CHAIN CORE PROTEIN VP3  
FT CHAIN CORE PROTEIN VP4  
FT CHAIN CORE PROTEIN VP5

DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol\_1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICIVIRUS.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
FT CHAIN 432 836 COAT PROTEIN VP1 (P1D).  
FT CHAIN 492 836 COAT PROTEIN VP1 (P1D).  
FT CHAIN 837 980 CORE PROTEIN P2A.  
FT CHAIN 981 1076 CORE PROTEIN P2B.  
FT CHAIN 1077 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1484 PROBABLE PROTEIN P3A.  
FT CHAIN 1485 1507 PROBABLE PROTEIN P3B.  
FT CHAIN 1495 1507 PROBABLE PROTEIN P3C.  
FT CHAIN 1508 1678 PROBABLE PROTEIN P3C.  
FT CHAIN 1679 2227 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;  
  
Query Match 100.0%; Score 107; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 4.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SHIECKPKYKELRLEVGVKQR 20  
DB 792 SHIECKPKYKELRLEVGVKQR 811  
  
RESULT 8  
POLG HPVAV STANDARD; PRT; 2227 AA.  
AC P13901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;  
AC Q81090; Q81091; Q81092; Q81093;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
P3D (EC 2.7.7.48)].  
DE Hepatitis A virus (strain MBB).  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12100;  
RN [1]\_TaxID=12100;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88045071; PubMed=2823500;  
RA Paul A.V., Tada H., der Helm K., Wessel T., Kiehn R., Wimmer E.,  
Deinhardt F.;  
RT "The entire nucleotide sequence of the genome of human hepatitis A  
virus (isolate MBB).";  
RL Virus Res. 8:153-171(1987).  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
{RNA}(N).  
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
each of which is composed of one copy each of proteins VP1, VP2,  
VP3, and VP4.  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -!- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT  
WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M20273; AAA45474.1; -.  
DR MEROPS; C03.005; -.

DR InterPro; IPR004004; Calici\_pol\_hel.  
DR InterPro; IPR009003; Cys\_ser\_trypsin.  
DR InterPro; IPR00605; RNA\_helicase.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol\_1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICIVIRUS.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
FT CHAIN 492 836 COAT PROTEIN VP1 (P1D).  
FT CHAIN 837 980 CORE PROTEIN P2A.  
FT CHAIN 981 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1496 PROBABLE PROTEIN P3A.  
FT CHAIN 1497 1519 PROBABLE PROTEIN P3B.  
FT CHAIN 1520 1738 PROBABLE PROTEIN P3C.  
FT CHAIN 1739 2227 RNA-DIRECTED POLYMERASE 3D.  
SQ SEQUENCE 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;  
  
Query Match 100.0%; Score 107; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 4.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SHIECKPKYKELRLEVGVKQR 20  
DB 792 SHIECKPKYKELRLEVGVKQR 811  
  
RESULT 9  
POLG HPVAV STANDARD; PRT; 808 AA.  
AC Q02381;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein  
P2A] (Fragment).  
DE Hepatitis A virus (strain GA76).  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=31706;  
RN [1]\_TaxID=31706;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92260183; PubMed=1316423;  
RA Khanna B., Spelbring J.E., Innis B.L., Robertson B.H.;  
RT "Characterization of a genetic variant of human hepatitis A virus.";  
RL J. Med. Virol. 36:118-124(1992).  
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
each of which is composed of one copy each of proteins VP1, VP2,  
VP3, and VP4.  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M66695; AAA45477.1; -.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
KW Polyprotein; Coat protein; Core protein.  
FT NON\_TER 1 1  
FT CHAIN <1 2 COAT PROTEIN VP4 (P1A).  
FT CHAIN 3 223 COAT PROTEIN VP2 (P1B).  
FT CHAIN 224 470 COAT PROTEIN VP3 (P1C).



```

FT CHAIN 471 770 COAT PROTEIN VP1 (P1D).
FT CHAIN 771 -808 CORE PROTEIN P2A.
FT NON_TER 808
SQ SEQUENCE 808 AA; 90632 MW; D80CE7B57A479C12 CRC64;

Query Match 88.88; Score 95; DB 1; Length 808;
Best Local Similarity 95.08; Pred.No. 1.4e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGKQK 20
|||||
Db 771 SHIECRKPYKELRLEVGKQK 790
|||||

RESULT 10
POLG HPAAVS STANDARD; PRT; 2230 AA.
P14553;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
OS Simian hepatitis A virus (strain AGM-27).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12102;
OX (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=91311420; PubMed=1649901;
RA Tsarev S.A., Emerson S.U., Balayan M.S., Ticehurst J.R.,
RA Purcell R.H.;
RA "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome
RT structure and growth in cell culture with other HAV strains.";
RT J. Gen. Virol. 72:1677-1683(1991).
RL [2]
RP SEQUENCE OF 1750-2164 FROM N.A.
RX MEDLINE=89323168; PubMed=2541023;
RA Balayan M.S., Kusov Y.Y., Andjaparidze A.G., Tsarev S.A.,
RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;
RA "Variations in genome fragments coding for RNA polymerase in human
RT and simian hepatitis A viruses.";
RT FEBS Lett. 247:425-428(1989).
CC - RNA(CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC - SUBUNIT: The virus capsid is composed of 60 icosahedral units.
CC each of which is composed of one copy each of proteins VP1, VP2,
CC VP3, and VP4.
CC - PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC - SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00924; BAA00766.1; -
CC EMBL; X15461; CAA33490.1; -
CC PIR; A30470; GNNYSA.
CC MEROPS; C03.005; -
DR InterPro; IPR004004; Calici_pol_hel.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR006050; RNA_helicase.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_P5vir.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.

```

QY 1 SHIECRKPYKELRLEVGVKQR 20  
 ||||:|||||||  
 Db 792 SHIE-KKPYKELRLEVGVKQR 810

RESULT 12	YAGA SCHPO	ID	YAGA SCHPO	STANDARD;	PRT;	420 AA.
AC	Q09873;					
DT	01-FEB-1996	(Rel. 33,	Created)			
DT	01-FEB-1996	(Rel. 33,	Last sequence update)			
DT	28-FEB-2003	(Rel. 41,	Last annotation update)			
DE	Hypothetical protein C12G12.10 in chromosome					
GN	SPAC12G12.10.					
OS	Schizosaccharomyces pombe (Fission yeast).					
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharo					
OC	Schizosaccharomycetales; Schizosaccharomycete					
OC	Schizosaccharomycetes.					
OX	NCBI_TaxID=4896;					
RN	[1]					
RP	SEQUENCE FROM N.A.					

MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sources J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall C., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neil S., Pearson D., Quayl M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Shelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Welltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaue V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Shrivatski L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Cerretti G.V., Ussery D., Barrill B.G., Nurse P.;  
RT "The genome sequence of *Schizosaccharomyces pombe*.";  
RL Nature 415:871-880(2002).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL curation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).

EMBL; Z66568; CA9A1505.1; -.  
PIR; S62541; S62541. -.  
GeneDB Spombe; SPAC12G12.10;  
InterPro; IPR001680; WD40.  
Pfam; PF00400; WD40; 1.  
SMART; SM00320; WD40; 3.  
Hypothetical protein.  
SEQUENCE 420 AA; 47525 MW

Query Match 43.0%; Score 46; DB 1; Length 420;  
Best Local Similarity 46.7%; Pred. No. 6.5;  
Matches 7; Conservative 5; Mismatches 3; Indels

Qy 4 ECRKPYKELRLEV GK 18

Db 376 DCSLPFKEMRVDDGK 390

RESULT 13

GC6P5	HUMAN	STANDARD;	PRT;	1819	AA.
ID	_GC6P5				
AC	Q36RT7; Q9BY91; Q9UCX3; Q9UCX4;				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Gamma-tubulin complex component 6 (GCP-6).				
GN	TUBGCP6 OR GCP6 OR KIA1659.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Verte				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hom				
OX	NCBI_TaxID=9606;				
RP	[1]				
RN	SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTER				
RX	MEDLINE=21551508; PubMed=1694571.				
RA	Murphy S.M., Preble A.M., Patel U.K., O'Conner				
RA	Moritz M., Agard D., Stults J.T., Stearns T.;				
RT	"GCP5 and GCP6: two new members of the human				
RL	Mol. Biol. Cell 12:3340-3352(2001).				

SEQUENCE FROM N.A.

Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,  
Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,  
Baggaley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,  
Burtil W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
Gilbert J.G.R., Goward M.E., Grafham D.W., Griffiths M.N.D., Hall C.,  
Hall R.E., Hall-Tamlyn G., Heathcott R.V., Ho S.S., Holmes S.,  
Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
Laird G.K., Langford C.P., Leversha M.A., Lloyd C., Lloyd D.M.,  
Martyn I.D., Mashregini-Mohammadi M., Matthews L.H., Mocan O.T.,  
Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,  
Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
Phillips S.H., Plumb R.M., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,  
Scott C.B., Sehra H.K., Skuce C.D., Smallie S., Smith M.L., Swann R.M.,  
Soderlund C., Spragon L., Steward C.A., Sulston J.E., Willet D.L.,  
Vaadin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
Shintani A., Shibuya K., Yoshiizaki Y., Aoki N., Mitsuyama S.,  
Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lac H.I.,  
Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,  
Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
Cordeas M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,  
Schuet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,  
Korf I., Bedell J.A., Hillier L.W., Marais E., Waterston R.,  
Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitsu S.,  
Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
Edelmann L., Kim U.J., Shiruya H., Simon M.I., Dumanski J.P.,  
Reyrad M., Kedra D., Serousi E., Fransson I., Tapia I., Bruder C.E.  
O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
Khan A.S., Lane L., Tilahun Y., Wright H.;  
"The DNA sequence of human chromosome 22,"  
Nature 402:489-495(1999).

[3]

SEQUENCE OF 33'-1819 FROM N.A. (ISOFORM 2).  
TISSUE=Brain;

```
RX MEDLINE=21156230; PubMed=11258795;
RA Hirosewa M., Nagase T., Murahehi Y., Kikuno R., Ohara O.;
RT "Identification of novel transcribed sequences on human chromosome 22
RL DNA Res. 8:1-9(2001).
CC -!- FUNCTION: Gamma-tubulin complex is necessary for microtubule
CC nucleation at the centrosome.
CC -!- SUBUNIT: Gamma-tubulin complex is composed of gamma-tubulin, GCP2,
CC GCP3, GCP4, GCP5 and GCP6.
CC -!- SUBCELLULAR LOCATION: Centrosome.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q96RT7-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q96RT7-2; Sequence=VSP_001624;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Belongs to the GCP family.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 1371 and 1758.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF272887; AAK82968.1; -
DR EMBL; AL022328; CAB63046.1; ALT_SEQ.
DR EMBL; AL022328; CAB63047.1; ALT_SEQ.
DR EMBL; AB051456; BAB33339.1; ALT_FRAME.
DR Genew; HGNC:18127; TUBGCP6.
DR GO; GO:0008274; C:gamma-tubulin ring complex; IDA.
DR GO; GO:0008017; F:microtubule binding; IDA.
DR GO; GO:0007020; F:microtubule nucleation; IDA.
DR InterPro; IPR007259; Spc97_Spc98.
DR Pfam; PF04130; Spc97_Spc98; 1.
DR DR Microtubule; Repeat; Alternative splicing.
DR DOMAIN 1027 1269 9 X 27 AA TANDEM REPEATS.
FT REPEAT 1027 1053 1.
FT REPEAT 1054 1080 2.
FT REPEAT 1081 1107 3.
FT REPEAT 1108 1134 4.
FT REPEAT 1135 1161 5.
FT REPEAT 1162 1188 6.
FT REPEAT 1189 1215 7.
FT REPEAT 1216 1242 8.
FT REPEAT 1243 1269 9.
FT VARSPLIC 1724 1757 Missing (in isoform 2).
FT FTID=VSP_001624.
FT CONFLICT 567 567 S -> L (IN REF. 3).
FT CONFLICT 1377 1377 A -> T (IN REF. 2 AND 3).
FT CONFLICT 1621 1621 L -> V (IN REF. 2).
SQ SEQUENCE 1819 AA; 200455 MW; 923576544D34594A CRC64;

Query Match 42.1%; Score 45; DB 1; Length 1819;
Best Local Similarity 50.0%; Pred. NO. 41;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 HIECRPKYKELRLEVGKQ 19
Db 653 HSSVSKEELRMEIAKQ 670

RESULT 14
CO3_RAT STANDARD; PRT; 1663 AA.
AC P01026;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
```

```
DB GN
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN SEQUENCE FROM N.A.
RP STRAIN=Wistar; TISSUE=Liver;
RX MEDLINE=90245672; PubMed=2336397;
RA Misumi Y., Sohda M., Ikehara Y.;
RT "Nucleotide and deduced amino acid sequence of rat complement C3.";
RL Nucleic Acids Res. 18:2178-2178(1990).
RN SEQUENCE OF 671-748.
RP MEDLINE=7906262; PubMed=309768;
RA Jacobs J.W., Rubin J.S., Hugli T.E., Bogardt R.A., Mariz I.K.,
RA Daniels J.S., Daughaday W.H., Bradshaw R.A.;
RT "Purification, characterization, and amino acid sequence of rat
anaphylatoxin (C3a)".
RL Biochemistry 17:5031-5038(1978).
RN SEQUENCE OF 1316-1595 FROM N.A.
RP MEDLINE=89380332; PubMed=2674144;
RA Sundstrom S.A., Komm B.S., Ponce-De-Leon H., Yi Z., Teuscher C.,
RA Lyttle C.R.;
RT "Estrogen regulation of tissue-specific expression of complement C3.";
RL J. Biol. Chem. 264:16941-16947(1989).
CC -!- FUNCTION: C3 plays a central role in the activation of the
CC complement system. Its processing by C3 convertase is the central
CC reaction in both classical and alternative complement pathways.
CC After activation C3b can bind covalently, via its reactive
CC thioester, to cell surface carbohydrates or immune aggregates.
CC -!- FUNCTION: Derived from proteolytic degradation of complement C3,
CC C3a anaphylatoxin is a mediator of local inflammatory process. It
CC induces the contraction of smooth muscle, increases vascular
CC permeability and causes histamine release from mast cells and
CC basophilic leukocytes.
CC -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA
CC CHAIN).
CC -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC -!- SIMILARITY: Contains 1 anaphylatoxin-like domain.
CC -!- SIMILARITY: Contains 1 NTR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52477; CAA36716.1; -
DR EMBL; M29866; AAA40837.1; ALT_SEQ.
DR PIR; S15764; C3PT.
DR PDB; 1QQF; 31-JUL-00.
DR PDB; 1QSU; 31-JUL-00.
DR InterPro; IPR002890; A2M_N.
DR InterPro; IPR009048; AM_receptor_bind.
DR InterPro; IPR000020; Anaphylatoxin.
DR InterPro; IPR001840; Anaphylatoxn.
DR InterPro; IPR001599; MacrogloblnA2.
DR InterPro; IPR001134; Netrin_C.
DR InterPro; IPR008930; Terp_cyc_toroid.
DR Pfam; PF00207; A2M; 1.
DR Pfam; PF01835; A2M_N; 1.
DR Pfam; PF01821; ANATO; 1.
DR Pfam; PF01759; NTR; 1.
DR PRINTS; PR00004; ANAPHYLATOXN.
DR ProDom; PD003264; Anaphylatoxin; 1.
```

```
DR SMART; SMO0104; ANATO; 1.
DR SMART; SMO0643; C345C; 1.
DR PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
DR PROSITE; PS01189; NTR; 1.
KW Complement pathway; Complement alternate pathway; Plasma;
KW Inflammatory response; Glycoprotein; Signal; 3D-structure;
FT THIOESTER BOND.
FT SIGNAL 1 24
FT CHAIN 25 1663
FT CHAIN 25 666
FT CHAIN 671 1663
FT PEPTIDE 671 748
FT CHAIN 749 1663
FT DOMAIN 693 728
FT DOMAIN 1518 1661
FT SITE 748 749
FT DISULFID 558 816
FT DISULFID 626 661
FT DISULFID 693 720
FT DISULFID 694 727
FT DISULFID 707 728
FT DISULFID 873 1513
FT DISULFID 1101 1158
FT DISULFID 1358 1489
FT DISULFID 1389 1458
FT DISULFID 1506 1511
FT DISULFID 1518 1590
FT DISULFID 1537 1661
FT DISULFID 1010 1013
FT CARBOHYD 939 939
FT CARBOHYD 1617 1617
FT CONFLICT 721 722
SQ SEQUENCE 1663 AA; 186460 MW; 2F87CCB143CDD4BC CRC64;

Query Match 40.7%; Score 43.5; DB 1; Length 1663;
Best Local Similarity 47.4%; Pred. No. 65;
Matches 9; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

QY 1 SHIECRKPKYKELRLEVGVKQ 19
Db 1586 SHVKCR---NALKLQKQKQ 1601

RESULT 15
NCAP_HRSV1
ID NCAP_HRSV1 STANDARD; PRT; 391 AA.
AC P24556;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE Nucleocapsid protein.
GN N.
OS Human respiratory syncytial virus (subgroup B / strain 18537).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
OX NCBI_TaxID=11251;
RN [1]
RP SEQUENCE FROM N.A. PubMed=2525176;
RX MEDLINE=89279331; Collins P.L.;
RA Johnson P.R.; Collins P.L.;
RT "The 1B (NS2), 1C (NS1) and N proteins of human respiratory syncytial
RT virus (RSV) of antigenic subgroups A and B: sequence conservation and
RT divergence within RSV genomic RNA.";
RL J. Gen. Virol. 70:1539-1547(1989).
CC -!- FUNCTION: Most abundant protein in the virion and an important
CC element conferring helical symmetry on the nucleoprotein core as
CC well as interacting with m protein during virion formation.
CC -!- SIMILARITY: Belongs to the paramyxoviruses nucleocapsid family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00736; BAA00637.1; -.
DR PIR; C32063; VHNZ3.
DR InterPro; IPR004930; Pneumo_ncap.
DR Pfam; PF03246; Pneumo_ncap; 1.
DR ProDom; PD006438; Pneumo_ncap; 1.
KW Nucleocapsid.
SQ SEQUENCE 391 AA; 43416 MW; 9A2EBB50103B2835 CRC64;

Query Match 40.2%; Score 43; DB 1; Length 391;
Best Local Similarity 56.2%; Pred. No. 18;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 IECKPKYKELRLEVGVK 18
Db 129 IESRKSYSKLLKEMGE 144

Search completed: March 15, 2004, 14:00:57
Job time : 6.64706 secs
```

OM protein - protein search, using sw model

Run on: March 15, 2004, 13:24:17 ; Search time 44.2353 Seconds  
(without alignments)  
127.748 Million cell updates/sec

Title: US-09-171-432A-39  
Perfect score: 107  
Sequence: 1 SHIECKPKYKELRLEVGKQR 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Genesecp1980s:\*  
2: Genesecp1990s:\*  
3: Genesecp2000s:\*  
4: Genesecp2001s:\*  
5: Genesecp2002s:\*  
6: Genesecp2003as:\*  
7: Genesecp2003bs:\*  
8: Genesecp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	107	100.0	20	2	AAW42922 Immunogen
2	107	100.0	21	4	AAB69439 Synthetic
3	107	100.0	366	1	AAP50230 Sequence
4	107	100.0	854	1	AAP50287 Hepatitis
5	107	100.0	993	1	AAP50116 Sequence
6	107	100.0	993	1	AAP50231 Sequence
7	107	100.0	1077	2	AAW95559 A partial
8	107	100.0	1091	2	AAW32426 Translate
9	107	100.0	2227	2	AAW5697 Attenuate
10	107	100.0	2227	2	AAW34074 Hepatitis
11	107	100.0	2227	3	AAB18609 Amino aci
12	107	100.0	2227	3	AAB18607 Amino aci
13	107	100.0	2227	3	AAB18608 Amino aci
14	107	100.0	2227	5	AAE19899 Hepatitis
15	107	100.0	2227	5	ABG31729 Attenuate
16	107	100.0	2227	5	ABG31727 Wild-type
17	107	100.0	2227	6	ABG31728 Hepatitis
18	107	100.0	2227	6	ABU08640 Attenuat
19	107	100.0	2227	6	ABU08641 Attenuat
20	107	100.0	2227	6	ABU08639 Wild type
21	107	100.0	2227	6	ABW00350 Hepatitis
22	91	85.0	2227	1	AAP60066 Sequence
23	84.5	79.0	839	2	AAR15629 Capsid re
24	67	62.6	20	2	AAW42923 Immunogen
25	67	62.6	21	4	AAB69440 Synthetic

26	45.5	42.5	178	4	AAU03662	AAU03662 Group B S
27	45.5	42.5	330	5	ABP30218	Abp30218 Streptoco
28	45.5	42.5	341	5	ABP25499	Abp25499 Streptoco
29	45	42.1	1273	5	ABP62888	Abp62888 Human pol
30	45	42.1	1437	6	ABP98857	Abp98857 Human str
31	45	42.1	1819	7	AD660309	Ad660309 Human Pro
32	44	41.1	71	3	AAG07411	Aag07411 Arabidops
33	44	41.1	88	3	AAE13254	Aae13254 Arabidops
34	44	41.1	211	4	AAE13255	Aae13255 Arabidops
35	44	41.1	346	2	AAE13255	Aae13255 Human tum
36	43	40.2	45	4	AAE13255	Aae13255 Human tum
37	43	40.2	45	5	AAE13255	Aae13255 Human tum
38	43	40.2	52	4	AAE13254	Aae13254 Human tum
39	43	40.2	52	5	AAE13254	Aae13254 Human tum
40	43	40.2	141	7	AAE13254	Aae13254 Human tum
41	43	40.2	186	6	ABM69585	Abm69585 Photornab
42	43	40.2	287	3	AAE13255	Aae13255 Arabidops
43	43	40.2	312	3	AAE13255	Aae13255 Arabidops
44	43	40.2	320	3	AAE13255	Aae13255 Arabidops
45	43	40.2	332	5	ABE91335	Abb91335 Herbicida

#### ALIGNMENTS

RESULT 1					
AAW42922					
ID	AAW42922 standard; peptide; 20 AA.				
XX					
AC	AAW42922;				
XX					
DT	28-APR-1998 (first entry)				
XX					
DE	Immunogenic Hepatitis A virus peptide YK-1315.				
XX					
XX	Immunogenic peptide; immunogenic epitope; P2A protein; immune response;				
KW	antibody.				
KW					
OS	Synthetic.				
OS	Hepatitis A virus.				
XX					
FN	WO9740147-A1.				
XX					
PD	30-OCT-1997.				
XX					
PF	18-APR-1997; 97WO-US006891.				
XX					
PR	19-APR-1996; 96US-0015644P.				
XX					
XX	(USSH ) US DEPT HEALTH & HUMAN SERVICES.				
XX	Fields HA, Khudyakov YE;				
XX	WPI; 1997-535831/49.				
XX					
PT	Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune				
PT	response to HAV in a mammal or to detect the presence of antibodies				
XX	against HAV in a mammal.				
XX					
PS	Claim 18; Page 112; 14Opp; English.				
XX					
XX	Peptides AAW42922-30 are immunogenic peptides corresponding to				
CC	immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are				
CC	substantially similar to a portion of the amino acid sequence of the p2A				
CC	protein of HAV corresponding to amino acids 792-811. The present peptide				
CC	is derived from amino acids 792-811, and has a reactivity of 54.2% with				
CC	acute sera. Compositions containing the peptides can be used to induce an				
CC	immune response to HAV in a mammal. The peptides can also be used to				
CC	detect the presence of antibodies against HAV in mammalian serum. The				
CC	peptides can also be used to make an antibody against HAV by				
XX	administering the peptide to a mammal				
SQ	Sequence 20 AA;				

Query Match 100.0%; Score 107; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20  
DB 1 SHIECRPKYKELRLEVGKQR 20

RESULT 2  
AAB69439  
ID AAB69439 standard; peptide; 21 AA.  
AC AAB69439;  
XX  
XX 20-APR-2001 (first entry)  
XX  
XX Synthetic HAV P2A peptide, SEQ ID NO: 39.  
DE  
XX Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
KW antigen; major structural capsid polypeptide; HAV antibody detection.  
KW  
XX Hepatitis A virus.  
OS Synthetic.  
OS  
XX WO200105924-A2.  
PN  
XX  
XX 25-JAN-2001.  
XX  
XX 14-JUL-2000; 2000WO-US019267.  
PF  
XX 15-JUL-1999; 99US-0144412P.  
PR  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
XX Fielde HA, Khudyakov YE;  
XX  
XX WPI; 2001-112681/12.  
DR  
XX  
XX Synthetic peptides used as antigen sources for enzyme immunoassays  
PT detecting anti-hepatitis A virus and as vaccines.  
PT  
XX Claim 13; Page 93; 130pp; English.  
XX  
XX The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IGM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy end  
CC of the peptides enhances the IGM antibody reactivity  
XX  
XX Sequence 21 AA;

Query Match 100.0%; Score 107; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20  
DB 1 SHIECRPKYKELRLEVGKQR 20

RESULT 3  
AAP50230  
ID AAP50230 standard; protein; 366 AA.  
XX  
XX AAP50230;  
AC  
XX 28-NOV-1991 (first entry)  
DT  
XX Sequence of hepatitis A virus (HAV) surface protein (VP-1).  
DE  
XX Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
KW diagnostic assay.  
KW  
XX Hepatitis A virus.  
OS  
XX EPI38704-A.  
PN  
XX 24-APR-1985.  
PD  
XX 09-OCT-1984; 84EP-00402025.  
PF  
XX 14-OCT-1983; 83US-00541836.  
PR  
XX 02-MAR-1984; 84US-00585942.  
PR  
XX (MERI ) MERCK & CO INC.  
PA  
XX Hughes JV, Scolnick EM, Tomassini JE;  
XX WPI; 1985-100818/17.  
XX DR N-ESDB; AAN50274.  
XX  
XX New hepatitis A virus surface protein - useful for binding to  
PT neutralising antibodies to the virus.  
PT  
XX Claim 21; Page 46-48; 49pp; English.  
PS  
XX VP1 is isolated by solubilisation of the intact virus in an aq. anionic  
CC surfactant and a reducing agent. The viral proteins are sepd. and the  
CC protein of molecular wt. 33000 daltons is sepd  
CC  
XX  
SQ Sequence 366 AA;

Query Match 100.0%; Score 107; DB 1; Length 366;  
Best Local Similarity 100.0%; Pred. No. 3.3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20  
DB 301 SHIECRPKYKELRLEVGKQR 320

RESULT 4  
AAP50287  
ID AAP50287 standard; protein; 854 AA.  
XX  
XX AAP50287;  
AC  
XX 25-MAR-2003 (revised)  
DT  
XX 30-NOV-1991 (first entry)  
DT  
XX Hepatitis A virus (HAV) peptide corresponding to the capsid protein  
DE region of poliovirus RNA.  
DE  
XX Hepatitis A virus assay; antigen; antibody.  
KW  
XX Hepatitis A virus.  
OS  
XX WO8501517-A.  
PN  
XX 11-APR-1985.  
PD  
XX 27-SEP-1984; 84WO-US001552.  
PF  
XX

PR 30-SEP-1983; 83US-00537911.  
 XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 PA Ticehurst JR, Baltimore D, Feinstein SM, Purcell RH;  
 PI Racanello VR;  
 XX WPI; 1985-098846/16.  
 DR N-PSDB; AAN50330.  
 XX New hepatitis A virus CDNA - useful in assays for the virus and for  
 PT prodn. of the viral antigen and antibodies to it.  
 XX Example; Fig 7; 60pp; English.  
 XX The inventors claim HAV cDNA and a method for producing it, whereby large  
 CC amts. can be obt'd. economically. The cDNA is useful in the assay for  
 CC detection of HAV quickly and easily and with high sensitivity and  
 CC specificity. The HAV cDNA is also used in the prodn. of HAV antigen or  
 CC antibodies to it. The antibodies may be monoclonal. (Updated on 25-MAR-  
 CC 2003 to correct PA field.)  
 XX SQ Sequence 854 AA;  
 Query Match 100.0%; Score 107; DB 1; Length 854;  
 Best Local Similarity 100.0%; Pred. No. 8e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SHIECKPKYKELRLVGKQR 20  
 DB 792 SHIECKPKYKELRLVGKQR 811  
 RESULT 5  
 AAP50116  
 ID AAP50116 standard; protein; 993 AA.  
 XX AAP50116;  
 AC AAP50116;  
 XX 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 30-SEP-1991 (first entry)  
 XX Sequence of Hepatitis A virus (HAV) immunogenic peptides VP-1, VP-2, VP-3  
 DE and VP-4.  
 XX Antigenic protein; immunogen; vaccine.  
 XX Hepatitis A virus; (strain CR326).  
 XX EP154587-A.  
 XX 11-SEP-1985.  
 XX 27-FEB-1985; 85EP-00400369.  
 XX 02-MAR-1984; 84US-00585819.  
 XX (MERI ) MERCK & CO INC.  
 XX Linemeyer DL, Menke JG, Rueben RG, Mitra SW;  
 XX WPI; 1985-224964/37.  
 DR N-PSDB; AAN50139.  
 XX New nucleotide sequences coding for hepatitis A virus antigens - useful  
 PT for eliciting normal immune response and in vaccines for protecting  
 PT against the virus.  
 XX Example; Page 11-17; 32pp; English.  
 XX Within the sequence in AAN50139 is encoded the information necessary to  
 CC make the antigenic proteins of HAV. The sequences encoding for the

CC structural proteins begin at base 403. The key sub-unit sequences within  
 CC VP-1, designated Sequences I,II,III,IV, and V, start, respectively at  
 CC 1882, 1963, 1999, 2146, 2347. Other nucleotide sequences which are  
 CC valuable as encoding antigenic proteins are the sequences from base 1749  
 CC to base 2722; from base 1487 to base 2980 and from base 1644 to base  
 CC 2722. The sequence from base 1749 to base 2722 is esp. valuable as a  
 CC vector for producing antigen protein. Sequences II-V are claimed. X in  
 CC AAP50116 denotes the translation of a stop codon. (Updated on 25-MAR-2003  
 CC to correct PI field.) (Updated on 24-OCT-2003 to standardise OS field)  
 XX SQ Sequence 993 AA;  
 Query Match 100.0%; Score 107; DB 1; Length 993;  
 Best Local Similarity 100.0%; Pred. No. 9.4e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SHIECKPKYKELRLVGKQR 20  
 DB 928 SHIECKPKYKELRLVGKQR 947  
 RESULT 6  
 AAP50231  
 ID AAP50231 standard; protein; 993 AA.  
 XX AAP50231;  
 AC AAP50231;  
 DT 28-NOV-1991 (first entry)  
 XX Sequence encoded by partial sequence of hepatitis A virus (HAV),  
 DE including surface protein (VP-1).  
 DE Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
 KW diagnostic assay.  
 XX Hepatitis A virus.  
 OS Key Location/Qualifiers  
 FH 628..993  
 FT Protein /note="claimed; X denotes translated stop codons and  
 FT unspecified triplets"  
 XX EP138704-A.  
 XX 24-APR-1985.  
 XX 09-OCT-1984; 84EP-00402025.  
 XX 14-OCT-1983; 83US-00541836.  
 PR 02-MAR-1984; 84US-00585942.  
 XX (MERI ) MERCK & CO INC.  
 XX Hughes JV, Scolnick EM, Tomassini JF;  
 PI WPI; 1985-100818/17.  
 DR N-PSDB; AAN50274.  
 XX New hepatitis A virus surface protein - useful for binding to  
 PT neutralising antibodies to the virus.  
 XX Disclosure; Page 17-23; 49pp; English.  
 XX VP1 is isolated by solubilisation of the intact virus in an aq. anionic  
 CC surfactant and a reducing agent. The viral proteins are sepd. and the  
 CC protein of molecular wt. 33000 daltons is sepd  
 XX SQ Sequence 993 AA;  
 Query Match 100.0%; Score 107; DB 1; Length 993;  
 Best Local Similarity 100.0%; Pred. No. 9.4e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 SHIECRKPYKELRLEVGVQR 20
Db 928 SHIECRKPYKELRLEVGVQR 947
|||||
|||||

RESULT 7
AAW95559
AAW95559 standard; protein; 1077 AA.
XX
XX
AAW95559;
AC
XX
XX
28-APR-1999 (first entry)
DT
XX
XX
A partial hepatitis A virus (HAV) protein.
DE
XX
XX
Hepatitis A virus protein; HAV; P2 region;
KW cell-culture-adapted HAV strain; infection; accelerated growth.
XX
XX
Hepatitis A virus.
OS
XX
XX
US5849562-A.
PN
XX
XX
15-DEC-1998.
PD
XX
XX
06-JUN-1995; 95US-00468926.
PF
XX
XX
30-SEP-1983; 83US-00537911.
PR
XX
XX
27-SEP-1984; 84US-00654942.
PR
XX
XX
06-OCT-1988; 88US-00256135.
PR
XX
XX
06-NOV-1991; 91US-00788252.
XX
XX
(USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX
XX
Emerson SU, Purcell RH;
PI
XX
XX
WFI; 1999-094412/08.
DR
XX
XX
N-PSDB; AAX01006.
XX
XX
Chimeric hepatitis A virus strains - with P2 region from cell-culture-
PT adapted strain in wild-type genome.
PT
XX
XX
Disclosure; Fig 7A-L; 36pp; English.
PS
XX
XX
The present sequence represents a partial hepatitis A virus (HAV)
CC protein. The specification describes a DNA construct consisting of a wild
CC -type HAV genome in which the P2 region is replaced by the P2 region from
CC a cell-culture-adapted HAV strain. The construct is used to demonstrate
CC that mutations in the P2 region of a cell-culture-adapted HAV strain are
CC sufficient for establishment of infection and accelerated growth in cell
CC culture
XX
XX
SQ Sequence 1077 AA;

Query Match 100.0%; Score 107; DB 2; Length 1077;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy 1 SHIECRKPYKELRLEVGVQR 20
Db 1015 SHIECRKPYKELRLEVGVQR 1034
|||||
|||||

RESULT 8
AAR32426
AAR32426 standard; protein; 1091 AA.
ID
XX
XX
AAR32426;
AC
XX
XX
27-AUG-2003 (revised)
DT
XX
XX
25-MAR-2003 (revised)
DT
XX
XX
17-DEC-2001 (revised)
DT
XX
XX
10-JUN-1993 (first entry)
DT
XX
XX

```



DT 24-OCT-2003 (revised)  
DT 27-AUG-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 15-AUG-1990 (first entry)  
XX  
DE Attenuated hepatitis A virus.  
XX  
KW Hepatitis A virus; vaccine; attenuated.  
XX  
OS Hepatitis A virus; strain HM-175.  
XX  
FH Key Location/Qualifiers  
FT Region 1. .23  
FT /label= VP4 = 1A  
FT Region 24. .245  
FT /label= VP2 = 1B  
FT Region 246. .491  
FT /label= VP3 = 1C  
FT Region 492. .791  
FT /label= VP1 = 1D  
FT Region 792. .980  
FT /label= 2A  
FT Region 981. .1087  
FT /label= 2B  
FT Region 1088. .1422  
FT /label= 2C  
FT Region 1423. .1496  
FT /label= 3A  
FT Region 1497. .1519  
FT /label= 3B = VPg  
FT Region 1520. .1738  
FT /label= 3C  
FT Region 1739. .2227  
FT /label= 3D  
XX  
PN US48994228-A.  
XX  
PD 16-JAN-1990.  
XX  
XX 12-JUL-1988; 88US-00217824.  
XX  
PR 19-SEP-1984; 84US-00652067.  
PR 09-SEP-1986; 86US-00905146.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
XX  
PI Purcell RH, Ticehurst JR, Cohen JL, Emerson SU, Feinstein SM;  
PI Daemer RJ, Gust ID;  
XX  
DR WPI; 1990-075557/10.  
DR N-PSDB; AAQ03512.  
XX  
PT Vaccine against hepatitis A virus infection - comprises novel attenuated  
PT hepatitis A virus strain.  
XX  
PS Claim 1; Fig 1; 18pp; English.  
XX  
XX The attenuated HAV is useful for inducing protective immunity against  
CC HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by  
CC several nucleotide changes distributed throughout the genome, is  
CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is  
CC suitable for use as an HAV vaccine. It is noted that not all the changes  
CC are necessary for attenuation and use as a vaccine. (Updated on 25-MAR-  
CC 2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PT field.)  
CC (Updated on 27-AUG-2003 to correct OS field.) (Updated on 24-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 2227 AA;  
Query Match 100.0%; Score 107; DB 2; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.2e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPKYKELRLVGVKQR 20  
Db |||||  
792 SHIECKPKYKELRLVGVKQR 811  
RESULT 10  
AAW34074  
ID AAW34074 standard; protein; 2227 AA.  
XX  
AC AAW34074;  
XX  
DT 17-OCT-2003 (revised)  
DT 27-APR-1998 (first entry)  
XX  
DE Hepatitis A virus HM-175 protein sequence.  
XX  
KW HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus; infection;  
KW vaccine.  
XX  
OS Hepatitis A virus; HM-175.  
XX  
FH Key Location/Qualifiers  
FT Protein 1. .23  
FT /label= VP4  
FT Protein 24. .245  
FT /label= VP2  
FT Protein 246. .491  
FT /label= VP3  
FT Protein 492. .791  
FT /label= VP1  
FT Protein 792. .980  
FT /label= 2A  
FT Protein 981. .1087  
FT /label= 2B  
FT Protein 1088. .1422  
FT /label= 2C  
FT Protein 1423. .1496  
FT /label= 3A  
FT Protein 1497. .1519  
FT /label= 3B  
FT Protein 1520. .1738  
FT /label= 3C  
FT Protein 1739. .2227  
FT /label= 3D  
XX  
PN WO9740166-A2.  
XX  
PD 30-OCT-1997.  
XX  
PF 18-APR-1997; 97WO-US006506.  
XX  
PR 19-APR-1996; 96US-0015642P.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Raychaudhuri G, Emerson SU, Purcell RH;  
XX  
DR WPI; 1997-535850/49.  
DR N-PSDB; AAT93023.  
XX  
PT Human attenuated HAV genome containing simian HAV 2C gene - useful as  
PT vaccines against HAV infection.  
XX  
PS Disclosure; Fig 13A-D; 66pp; English.  
XX  
CC This protein sequence is encoded by the human hepatitis A virus (HAV) HM-  
CC 175 wild-type genome (see AAT93023). Attenuated strain HAV/7 is obtained  
CC by passage of HM-175 in African Green Monkey kidney cells. A claimed DNA  
CC construct (I) comprises a genome of HAV, where the genome is a human  
CC attenuated HAV genome in which a region of the 2C gene has been replaced  
CC (see AAT93024). The region of the 2C gene from AGM-27 contained in the  
CC construct preferably encodes amino acids 120-328 of the 2C protein, amino

CC acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA transcript of (1); (2) a cell transfected with (1) or the RNA transcript of (1); (3) a HAV genome as above; (4) antibodies to the HAV of (3); and (5) a host cell containing the HAV of (3). (1) or its RNA transcript, can be used as a vaccine for preventing HAV in a mammal. (1) or the RNA transcript can also be used to stimulate the production of protective antibodies in the mammal. (Updated on 17-OCT-2003 to standardise OS field)

XX Sequence 2227 AA;

Query Match 100.0%; Score 107; DB 2; Length 2227;

Best Local Similarity 100.0%; Pred. No. 2.2e-07; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0;

QY 1 SHIECRPKYKELRLEVGKQR 20

|||||

DB 792 SHIECRPKYKELRLEVGKQR 811

RESULT 11

AAAB18609

ID AAB18609 standard; protein; 2227 AA.

XX AAB18609;

XX 15-JAN-2001 (first entry)

DE Amino acid sequence of live attenuated Hepatitis A virus 4380.

XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;

KW HAV 4380.

OS Hepatitis A virus.

XX US6113912-A.

XX 05-SEP-2000.

XX 07-JUN-1995; 95US-00475886.

XX 18-SEP-1992; 92US-00947338.

XX 17-SEP-1993; 93WO-US008610.

XX 17-APR-1995; 95US-00397232.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;

XX WPI; 2000-586464/55.

XX N-PSDB; AAA75478.

XX Novel live hepatitis A virus adapted to growth in human fibroblast cell line useful as vaccine for protecting humans against hepatitis A virus infection, has modified genome compared to wild type.

XX Disclosure; Col 93-104; 72pp; English.

XX The present sequence is derived from a live attenuated hepatitis A virus (HAV) of the invention, designated HAV 4380. The sequence is produced by modifying wild type HAV strain HM-174. The HAV of the invention are adapted to growth in the human fibroblast-like cell line MRC-5. The HAV is able to propagate in MRC-5 cells and retain appropriate attenuation. It is useful as a live vaccine for prophylaxis of hepatitis A in humans and other primates

XX Sequence 2227 AA;

Query Match 100.0%; Score 107; DB 3; Length 2227;

Best Local Similarity 100.0%; Pred. No. 2.2e-07; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20

|||||

Db 792 SHIECRPKYKELRLEVGKQR 811

RESULT 12

AAAB18607

ID AAB18607 standard; protein; 2227 AA.

XX AAB18607;

XX 15-JAN-2001 (first entry)

DE Amino acid sequence of wild type Hepatitis A virus strain HM-175.

XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.

OS Hepatitis A virus.

XX US6113912-A.

XX 05-SEP-2000.

XX 07-JUN-1995; 95US-00475886.

XX 18-SEP-1992; 92US-00947338.

XX 17-SEP-1993; 93WO-US008610.

XX 17-APR-1995; 95US-00397232.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;

XX WPI; 2000-586464/55.

XX N-PSDB; AAA75476.

XX Novel live hepatitis A virus adapted to growth in human fibroblast cell line useful as vaccine for protecting humans against hepatitis A virus infection, has modified genome compared to wild type.

XX Disclosure; Fig 6A-K; 72pp; English.

XX The present sequence is derived from a wild type hepatitis A virus (HAV) strain HM-174. The sequence is modified to produce HAV which are adapted to growth in the human fibroblast-like cell line MRC-5. The HAV is able to propagate in MRC-5 cells and retain appropriate attenuation. It is useful as a live vaccine for prophylaxis of hepatitis A in humans and other primates

XX Sequence 2227 AA;

Query Match 100.0%; Score 107; DB 3; Length 2227;

Best Local Similarity 100.0%; Pred. No. 2.2e-07; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20

|||||

Db 792 SHIECRPKYKELRLEVGKQR 811

RESULT 13

AAAB18608

ID AAB18608 standard; protein; 2227 AA.

XX AAB18608;

XX 15-JAN-2001 (first entry)

DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.

XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;

KW P-35 virus.

OS Hepatitis A virus.

XX US6113912-A.  
XX 05-SEP-2000.  
XX 07-JUN-1995; 95US-00475886.  
XX 18-SEP-1992; 92US-00947338.  
XX 17-SEP-1993; 93WO-US008610.  
XX 17-APR-1995; 95US-00397232.  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
XX N-PSDB; AAA75477.  
XX WPI; 2000-586464/55.  
XX N-PSDB; AAA75477.  
XX Novel live hepatitis A virus adapted to growth in human fibroblast cell  
XX line useful as vaccine for protecting humans against hepatitis A virus  
XX infection, has modified genome compared to wild type.  
XX Disclosure; Col 67-78; 72pp; English.  
XX The present sequence is derived from passage 35 of a wild type hepatitis  
XX A virus (HAV) strain HM-174. The resulting virus is designated P-35  
XX virus. The sequence is modified to produce HAV which are adapted to  
XX growth in the human fibroblast-like cell line MRC-5. The HAV is able to  
XX propagate in MRC-5 cells and retain appropriate attenuation. It is useful  
XX as a live vaccine for prophylaxis of hepatitis A in humans and other  
XX primates.  
XX Sequence 2227 AA;  
Query Match 100.0%; Score 107; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.2e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECKPKYKELRLEVGKOR 20  
DB 792 SHIECKPKYKELRLEVGKOR 811  
RESULT 14  
AAE19899  
ID AAE19899 standard; protein; 2227 AA.  
XX AC AAE19899;  
XX DT 18-JUN-2002 (first entry)  
XX DE Hepatitis A virus (HAV) protein.  
XX KW Hepatitis A virus; HAV; infection; virucide; fungicide; antibacterial;  
XX KW cytostatic; immunostimulant; vaccine; ribavirin; immune response; cancer.  
XX OS Hepatitis A virus.  
XX PN WO200213855-A2.  
XX PD 21-FEB-2002.  
XX PF 15-AUG-2001; 2001WO-IB001808.  
XX PR 17-AUG-2000; 2000US-0225767P.  
XX PR 29-AUG-2000; 2000US-0229175P.  
XX PR 03-NOV-2000; 2000US-00705547.  
XX (TRIP-) TRIPEP AB.  
XX Sallberg M, Hultgren C;  
XX WPI; 2002-241837/29.  
XX N-PSDB; AAD31766.  
XX Vaccine compositions for treating and preventing disease, preferably  
XX hepatitis C virus infection, comprises ribavirin and antigen that has  
XX epitope present in hepatitis C virus.  
XX Claim 11; Page 82-87; 120pp; English.  
XX The invention relates to a composition comprising ribavirin and an  
XX antigen preferably non structural 3 protein (NS3)/4A fragment of  
XX hepatitis C virus (HCV) genome or a peptide or nucleic acid of HCV  
XX sequence. The composition is useful for enhancing an immune response to a  
XX hepatitis C antigen in humans, domestic, sport or pet species and as  
XX vaccines for treating and preventing HCV infections. The composition is  
XX also useful for treating viral, bacterial, fungal diseases and cancer.  
XX The present sequence is hepatitis A virus (HAV) protein  
XX Sequence 2227 AA;  
Query Match 100.0%; Score 107; DB 5; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.2e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECKPKYKELRLEVGKOR 20  
DB 792 SHIECKPKYKELRLEVGKOR 811  
RESULT 15  
ABG31729  
ID ABG31729 standard; protein; 2227 AA.  
XX AC ABG31729;  
XX DT 29-AUG-2003 (revised)  
XX DT 29-NOV-2002 (first entry)  
XX DE Attenuated Hepatitis A virus (HAV) 4380 polypeptide.  
XX KW Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast; virucide;  
XX KW HAV 4380.  
XX OS Hepatitis A virus; strain HM-175.  
XX PN US6423318-B1.  
XX PD 23-JUL-2002.  
XX PF 31-AUG-2000; 2000US-00653499.  
XX PR 17-SEP-1993; 93WO-US008610.  
XX PR 17-APR-1995; 95US-00397232.  
XX PR 07-JUN-1995; 95US-00475886.  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
XX WPI; 2002-680946/73.  
XX N-PSDB; ABS52789.  
XX New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth  
XX in MRC-5 cells, useful for preparing a vaccine against HAV infection.  
XX Disclosure; Col 93-104; 71pp; English.  
XX The invention relates to a polynucleotide which encodes a hepatitis A  
XX virus (HAV) adapted to growth in MRC-5 cells (human fibroblast-like cell  
XX line). The polynucleotide is useful for preparing a vaccine against  
XX hepatitis A virus infection. This sequence represents an attenuated  
XX hepatitis A virus 4830 polypeptide. (Updated on 29-AUG-2003 to  
XX standardise OS field)

PN US6113912-A.  
XX 05-SEP-2000.  
XX 07-JUN-1995; 95US-00475886.  
XX 18-SEP-1992; 92US-00947338.  
XX 17-SEP-1993; 93WO-US008610.  
XX 17-APR-1995; 95US-00397232.  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
XX N-PSDB; AAA75477.  
XX WPI; 2000-586464/55.  
XX N-PSDB; AAA75477.  
XX Novel live hepatitis A virus adapted to growth in human fibroblast cell  
XX line useful as vaccine for protecting humans against hepatitis A virus  
XX infection, has modified genome compared to wild type.  
XX Disclosure; Col 67-78; 72pp; English.  
XX The present sequence is derived from passage 35 of a wild type hepatitis  
XX A virus (HAV) strain HM-174. The resulting virus is designated P-35  
XX virus. The sequence is modified to produce HAV which are adapted to  
XX growth in the human fibroblast-like cell line MRC-5. The HAV is able to  
XX propagate in MRC-5 cells and retain appropriate attenuation. It is useful  
XX as a live vaccine for prophylaxis of hepatitis A in humans and other  
XX primates.  
XX Sequence 2227 AA;  
Query Match 100.0%; Score 107; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.2e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECKPKYKELRLEVGKOR 20  
DB 792 SHIECKPKYKELRLEVGKOR 811  
RESULT 14  
AAE19899  
ID AAE19899 standard; protein; 2227 AA.  
XX AC AAE19899;  
XX DT 18-JUN-2002 (first entry)  
XX DE Hepatitis A virus (HAV) protein.  
XX KW Hepatitis A virus; HAV; infection; virucide; fungicide; antibacterial;  
XX KW cytostatic; immunostimulant; vaccine; ribavirin; immune response; cancer.  
XX OS Hepatitis A virus.  
XX PN WO200213855-A2.  
XX PD 21-FEB-2002.  
XX PF 15-AUG-2001; 2001WO-IB001808.  
XX PR 17-AUG-2000; 2000US-0225767P.  
XX PR 29-AUG-2000; 2000US-0229175P.  
XX PR 03-NOV-2000; 2000US-00705547.  
XX (TRIP-) TRIPEP AB.  
XX Sallberg M, Hultgren C;  
XX WPI; 2002-241837/29.  
XX N-PSDB; AAD31766.  
XX Vaccine compositions for treating and preventing disease, preferably  
XX hepatitis C virus infection, comprises ribavirin and antigen that has  
XX epitope present in hepatitis C virus.  
XX Claim 11; Page 82-87; 120pp; English.  
XX The invention relates to a composition comprising ribavirin and an  
XX antigen preferably non structural 3 protein (NS3)/4A fragment of HCV  
XX genome or a peptide or nucleic acid of HCV sequence. The composition is  
XX useful for enhancing an immune response to a hepatitis C antigen in  
XX humans, domestic, sport or pet species and as vaccines for treating and  
XX preventing HCV infections. The composition is also useful for treating  
XX viral, bacterial, fungal diseases and cancer. The present sequence is  
XX hepatitis A virus (HAV) protein  
XX Sequence 2227 AA;  
Query Match 100.0%; Score 107; DB 5; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.2e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECKPKYKELRLEVGKOR 20  
DB 792 SHIECKPKYKELRLEVGKOR 811  
RESULT 15  
ABG31729  
ID ABG31729 standard; protein; 2227 AA.  
XX AC ABG31729;  
XX DT 29-AUG-2003 (revised)  
XX DT 29-NOV-2002 (first entry)  
XX DE Attenuated Hepatitis A virus (HAV) 4380 polypeptide.  
XX KW Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast; virucide;  
XX KW HAV 4380.  
XX OS Hepatitis A virus; strain HM-175.  
XX PN US6423318-B1.  
XX PD 23-JUL-2002.  
XX PF 31-AUG-2000; 2000US-00653499.  
XX PR 17-SEP-1993; 93WO-US008610.  
XX PR 17-APR-1995; 95US-00397232.  
XX PR 07-JUN-1995; 95US-00475886.  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
XX WPI; 2002-680946/73.  
XX N-PSDB; ABS52789.  
XX New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth  
XX in MRC-5 cells, useful for preparing a vaccine against HAV infection.  
XX Disclosure; Col 93-104; 71pp; English.  
XX The invention relates to a polynucleotide which encodes a hepatitis A  
XX virus (HAV) adapted to growth in MRC-5 cells (human fibroblast-like cell  
XX line). The polynucleotide is useful for preparing a vaccine against  
XX hepatitis A virus infection. This sequence represents an attenuated  
XX hepatitis A virus 4830 polypeptide. (Updated on 29-AUG-2003 to  
XX standardise OS field)

SQ Sequence 2227 AA;

Query Match 100.0%; Score 107; DB 5; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.2e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGKQR 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 792 SHIECRKPYKELRLEVGKQR 811

Search completed: March 15, 2004, 13:59:56  
Job time : 45.2353 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2004, 13:15:47 ; Search time 11.2941 Seconds  
(without alignments)  
91.421 Million cell updates/sec

Title: US-09-171-432A-39

Perfect score: 107

Sequence: 1 SHIECKPKYKELRLEVGKQR 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A COMB.pap:\*
- 2: /cgn2\_6/ptodata/2/iaa/5B COMB.pap:\*
- 3: /cgn2\_6/ptodata/2/iaa/6A COMB.pap:\*
- 4: /cgn2\_6/ptodata/2/iaa/6B COMB.pap:\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pap:\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	1091	6	Patent No. 5516630-2
2	107	100.0	2227	3	US-08-475-886-2
3	107	100.0	2227	3	US-08-475-886-4
4	107	100.0	2227	3	US-08-475-886-6
5	107	100.0	2227	3	US-08-397-232-2
6	107	100.0	2227	3	US-08-397-232-4
7	107	100.0	2227	3	US-09-171-387-2
8	107	100.0	2227	4	US-09-653-499-2
9	107	100.0	2227	4	US-09-653-499-4
10	107	100.0	2227	4	US-09-653-499-6
11	107	100.0	2227	4	US-10-104-968-12
12	107	100.0	2227	4	US-10-135-988-2
13	107	100.0	2227	4	US-10-135-988-4
14	107	100.0	2227	4	US-10-135-988-6
15	84.5	79.0	839	1	US-08-087-016-2
16	43	40.2	1127	3	US-09-150-460B-11
17	42.5	39.7	5405	3	US-08-718-388-9
18	42	39.3	444	4	US-09-252-991A-32415
19	42	39.3	472	4	US-09-328-352-6296
20	42	39.3	977	4	US-09-543-681A-6386
21	41	38.3	114	4	US-09-489-847-321
22	41	38.3	153	4	US-09-533-029-96
23	41	38.3	391	5	PCT-US91-08177-3
24	41	38.3	534	4	US-09-252-991A-32086
25	41	38.3	607	4	US-09-907-794A-190
26	41	38.3	607	4	US-09-905-125A-190
27	41	38.3	607	4	US-09-902-775A-190

```

28 41 38.3 843 4 US-09-491-356C-20 Sequence 20, Appl
29 41 38.3 4544 1 US-08-469-486-52 Sequence 52, Appl
30 41 38.3 4544 2 US-08-469-658-52 Sequence 52, Appl
31 40 37.4 110 4 US-09-341-461-21 Sequence 21, Appl
32 40 37.4 414 4 US-09-252-991A-26659 Sequence 26659, A
33 39 36.4 24 1 US-08-406-347A-13 Sequence 13, Appl
34 39 36.4 154 4 US-09-198-452A-530 Sequence 530, App
35 39 36.4 238 4 US-09-257-179-80 Sequence 80, Appl
36 39 36.4 305 4 US-09-635-872A-3 Sequence 3, Appl
37 39 36.4 305 4 US-09-636-077A-3 Sequence 3, Appl
38 39 36.4 305 4 US-09-636-060C-3 Sequence 3, Appl
39 39 36.4 305 4 US-09-986-552-3 Sequence 3, Appl
40 39 36.4 344 3 US-09-393-554-2 Sequence 2, Appl
41 39 36.4 495 4 US-09-252-991A-25802 Patent No. 5223391
42 39 36.4 520 6 5223391-5 Patent No. 5223391
43 39 36.4 892 4 US-09-585-858-16 Sequence 16, Appl
44 39 36.4 1045 4 US-09-489-039A-8589 Sequence 8589, Ap
45 39 36.4 1124 4 US-09-252-991A-26810 Sequence 26810, A

```

#### ALIGNMENTS

```

RESULT 1
5516630-2
; Patent No. 5516630
; APPLICANT: TICHEURST, JOHN R.; BALTIMORE, DAVID; FEINSTONE,
; STEPHEN M.; PURCELL, ROBERT H.; RACANELLO, VINCENT R.;
; BAROUDY, BAHTIGE M.
; TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/788,262
; FILING DATE: 06-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 256,135
; FILING DATE: 06-OCT-1988
; APPLICATION NUMBER: 654,942
; FILING DATE: 27-SEP-1984
; APPLICATION NUMBER: 537,911
; FILING DATE: 30-SEP-1983
; SEQ ID NO:2:
; LENGTH: 1091
5516630-2

```

```

Query Match 100.0%; Score 107; DB 6; Length 1091;
Best Local Similarity 100.0%; Pred. No. 8.4e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 SHIECKPKYKELRLEVGKQR 20
Db 1029 SHIECKPKYKELRLEVGKQR 1048

```

```

RESULT 2
US-08-475-886-2
; Sequence 2, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1

```

```

; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-475-886-2

Query Match
Best Local Similarity 100.0%; Score 107; DB 3; Length 2227;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20
Db 792 SHIECRPKYKELRLEVGKQR 811

RESULT 3
US-08-475-886-4
; Sequence 4, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-08-475-886-4

Query Match
Best Local Similarity 100.0%; Score 107; DB 3; Length 2227;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20
Db 792 SHIECRPKYKELRLEVGKQR 811

RESULT 4
US-08-475-886-6
; Sequence 6, Application US/08475886A
; Patent No. 6113912
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-475-886-6

Query Match
Best Local Similarity 100.0%; Score 107; DB 3; Length 2227;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20
Db 792 SHIECRPKYKELRLEVGKQR 811

RESULT 5
US-08-397-232-2
; Sequence 2, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-08-397-232-2

Query Match
Best Local Similarity 100.0%; Score 107; DB 3; Length 2227;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20
Db 792 SHIECRPKYKELRLEVGKQR 811

RESULT 6
US-08-397-232-4
; Sequence 4, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US1
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-04-17
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: PCT/US93/08610
; EARLIER FILING DATE: 1993-09-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-4

Query Match
Best Local Similarity 100.0%; Score 107; DB 3; Length 2227;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVGKQR 20
Db 792 SHIECRPKYKELRLEVGKQR 811

RESULT 7
US-08-397-232-6
; Sequence 6, Application US/08397232A
; Patent No. 6180110
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/397,232A
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-08-397-232-6

```

```
QY 1 SHIECKPKYKELRLVGVKQR 20
Db 792 SHIECKPKYKELRLVGVKQR 811

RESULT 7
US-09-171-387-2
; Sequence 2, Application US/09171387
; Patent No. 6280734
; GENERAL INFORMATION:
; APPLICANT: RAYCHAUDHURI, GOPA;
; EMERSON, SUZANNE, U.;
; PURCELL, ROBERT, H.
; TITLE OF INVENTION: SIMIAN-HUMAN HAV
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/171,387
; FILING DATE: 24-Mar-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/06506
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US60/015,642
; FILING DATE: 19-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: William S. Feiler
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4229US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 751-6849
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 2
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2227 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-171-387-2

Query Match 100.0%; Score 107; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPKYKELRLVGVKQR 20
Db 792 SHIECKPKYKELRLVGVKQR 811

RESULT 8
US-09-653-499-2
; Sequence 2, Application US/09653499
; Patent No. 6423318
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; EMERSON, SUZANNE U
; PURCELL, ROBERT H
; PURCELL, ROBERT H
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/09/653,499
; CURRENT FILING DATE: 2000-08-31
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 08/475,886
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-09-653-499-2

Query Match 100.0%; Score 107; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPKYKELRLVGVKQR 20
Db 792 SHIECKPKYKELRLVGVKQR 811

RESULT 9
US-09-653-499-4
; Sequence 4, Application US/09653499
; Patent No. 6423318
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; EMERSON, SUZANNE U
; PURCELL, ROBERT H
; PURCELL, ROBERT H
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/09/653,499
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 08/475,886
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-09-653-499-4

Query Match 100.0%; Score 107; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPKYKELRLVGVKQR 20
Db 792 SHIECKPKYKELRLVGVKQR 811

RESULT 10
US-09-653-499-6
; Sequence 6, Application US/09653499
; Patent No. 6423318
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; EMERSON, SUZANNE U
; PURCELL, ROBERT H
; PURCELL, ROBERT H
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/09/653,499
; CURRENT FILING DATE: 2000-08-31
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 08/475,886
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-09-653-499-6

Query Match 100.0%; Score 107; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPKYKELRLVGVKQR 20
Db 792 SHIECKPKYKELRLVGVKQR 811

RESULT 11
US-09-171-387-2
; Sequence 2, Application US/09171387
; Patent No. 6280734
; GENERAL INFORMATION:
; APPLICANT: RAYCHAUDHURI, GOPA;
; EMERSON, SUZANNE, U.;
; PURCELL, ROBERT, H.
; TITLE OF INVENTION: SIMIAN-HUMAN HAV
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/171,387
; FILING DATE: 24-Mar-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/06506
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US60/015,642
; FILING DATE: 19-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: William S. Feiler
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4229US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 751-6849
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 2
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2227 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-171-387-2

Query Match 100.0%; Score 107; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPKYKELRLVGVKQR 20
Db 792 SHIECKPKYKELRLVGVKQR 811

RESULT 12
US-09-653-499-2
; Sequence 2, Application US/09653499
; Patent No. 6423318
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; EMERSON, SUZANNE U
; PURCELL, ROBERT H
; PURCELL, ROBERT H
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/09/653,499
; CURRENT FILING DATE: 2000-08-31
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 08/475,886
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-09-653-499-2

Query Match 100.0%; Score 107; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPKYKELRLVGVKQR 20
Db 792 SHIECKPKYKELRLVGVKQR 811

RESULT 13
US-09-653-499-4
; Sequence 4, Application US/09653499
; Patent No. 6423318
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; EMERSON, SUZANNE U
; PURCELL, ROBERT H
; PURCELL, ROBERT H
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/09/653,499
; CURRENT FILING DATE: 2000-08-31
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 08/475,886
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-09-653-499-4

Query Match 100.0%; Score 107; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPKYKELRLVGVKQR 20
Db 792 SHIECKPKYKELRLVGVKQR 811

RESULT 14
US-09-653-499-6
; Sequence 6, Application US/09653499
; Patent No. 6423318
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; EMERSON, SUZANNE U
; PURCELL, ROBERT H
; PURCELL, ROBERT H
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/09/653,499
; CURRENT FILING DATE: 2000-08-31
; CURRENT FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 08/475,886
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-09-653-499-6

Query Match 100.0%; Score 107; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPKYKELRLVGVKQR 20
Db 792 SHIECKPKYKELRLVGVKQR 811
```

; PRIOR APPLICATION NUMBER: 08/475,886  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; SOFTWARE: Patent in Ver. 2.1  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-10-135-988-2

Query Match 100.0%; Score 107; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGKQR 20  
DB 792 SHIECRKPYKELRLEVGKQR 811

RESULT 13

US-10-135-988-4  
; Sequence 4, Application US/10135988  
; Patent No. 6680060  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175  
US-10-135-988-4

Query Match 100.0%; Score 107; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGKQR 20  
DB 792 SHIECRKPYKELRLEVGKQR 811

RESULT 14

US-10-135-988-6  
; Sequence 6, Application US/10135988  
; Patent No. 6680060  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 6

; PRIOR APPLICATION NUMBER: 08/475,886  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; SOFTWARE: Patent in Ver. 2.1  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 6  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-09-653-499-6

Query Match 100.0%; Score 107; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGKQR 20  
DB 792 SHIECRKPYKELRLEVGKQR 811

RESULT 11

US-10-104-966-12  
; Sequence 12, Application US/10104966  
; Patent No. 6680059  
; GENERAL INFORMATION:  
; APPLICANT: Matti Sallberg  
; APPLICANT: Catharina Hultgren  
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND  
; FILE REFERENCE: TRIPEP.23AUSC1  
; CURRENT APPLICATION NUMBER: US/10/104,966  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/705,547  
; PRIOR FILING DATE: 2000-11-03  
; PRIOR APPLICATION NUMBER: 60/229,175  
; PRIOR FILING DATE: 2000-08-29  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Hepatitis A virus sequence  
US-10-104-966-12

Query Match 100.0%; Score 107; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPYKELRLEVGKQR 20  
DB 792 SHIECRKPYKELRLEVGKQR 811

RESULT 12

US-10-135-988-2  
; Sequence 2, Application US/10135988  
; Patent No. 6680060  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18



; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-10-135-988-6  
Query Match 100.0%; Score 107; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECKPKYKELRLEVGVKQR 20  
DB 792 SHIECKPKYKELRLEVGVKQR 811  
|||||:|||||

RESULT 15  
US-08-087-016-2  
; Sequence 2, Application US/08087016  
; Patent No. 5430135  
; GENERAL INFORMATION:  
; APPLICANT: NAINAN, OVANA V.  
; APPLICANT: MARGOLIS, HAROLD S.  
; APPLICANT: ROBERTSON, BETTY H.  
; APPLICANT: BRINTON, MARGO H.  
; APPLICANT: EBERT, JAMES W.  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1615 L Street N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/087,016  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/678,828  
; FILING DATE: 03-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26,581  
; REFERENCE/DOCKET NUMBER: WTS/5683/83834/SRL  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 839 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-087-016-2

Query Match 79.0%; Score 84.5; DB 1; Length 839;  
Best Local Similarity 90.0%; Pred. No. 4.4e-06;  
Matches 18; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 1 SHIECKPKYKELRLEVGVKQR 20  
DB 792 SHIECKPKYKELRLEVGVKQR 810  
|||||:|||||

Search completed: March 15, 2004, 13:26:03  
Job time : 12.2941 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2004, 13:23:07 ; Search time 23.1765 Seconds  
(without alignments)  
182.213 Million cell updates/sec

Title: US-09-171-432A-39  
Perfect score: 107  
Sequence: 1 SHIECRKPKYKELRLEVGKQR 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	107	100.0	20	10	US-09-171-432A-39
2	107	100.0	352	14	US-10-272-459-45
3	107	100.0	836	14	US-10-272-459-40
4	107	100.0	980	14	US-10-272-459-41
5	107	100.0	2227	9	US-09-929-953-12
6	107	100.0	2227	13	US-10-104-966-12
7	107	100.0	2227	13	US-10-135-988-2
8	107	100.0	2227	13	US-10-135-988-4
9	107	100.0	2227	13	US-10-135-988-6
10	67	62.6	20	10	US-09-171-432A-40
11	45.5	42.5	178	14	US-10-091-007-124
12	43	40.2	45	9	US-09-728-912-13
13	43	40.2	52	9	US-09-728-912-12
14	43	40.2	413	9	US-09-728-912-2
15	42.5	39.7	5405	9	US-09-922-217-1116

16	42.5	39.7	5405	13	US-10-025-380-1116
17	42	39.3	1087	15	US-10-369-493-6801
18	42	39.3	1289	9	US-09-932-145-11
19	42	39.3	1289	15	US-10-395-812-11
20	41	38.3	22	14	US-10-173-461-21
21	41	38.3	39	14	US-10-173-461-65
22	41	38.3	71	9	US-09-864-761-43211
23	41	38.3	80	9	US-09-804-156-38
24	41	38.3	80	13	US-10-067-761-38
25	41	38.3	80	14	US-10-319-519-38
26	41	38.3	153	10	US-09-533-029-96
27	41	38.3	153	10	US-09-934-455-40
28	41	38.3	153	14	US-10-286-264-108
29	41	38.3	153	14	US-10-278-536-126
30	41	38.3	153	15	US-10-225-068-38
31	41	38.3	153	15	US-10-374-780A-2208
32	41	38.3	160	15	US-10-330-051A-44
33	41	38.3	254	9	US-09-820-893-94
34	41	38.3	264	14	US-10-173-461-2
35	41	38.3	315	9	US-09-820-893-59
36	41	38.3	324	9	US-09-820-893-96
37	41	38.3	327	15	US-10-369-493-3173
38	41	38.3	364	15	US-10-369-493-21445
39	41	38.3	492	9	US-09-978-295A-7
40	41	38.3	492	9	US-09-978-697-7
41	41	38.3	492	9	US-09-978-192A-7
42	41	38.3	492	9	US-09-999-832A-7
43	41	38.3	492	10	US-09-978-189-7
44	41	38.3	492	10	US-09-978-608A-7
45	41	38.3	492	10	US-09-978-585A-7

ALIGNMENTS

RESULT 1  
US-09-171-432A-39

; Sequence 39, Application US/09171432A  
; Publication No. US20030187184A1  
; GENERAL INFORMATION:  
; APPLICANT: Fields, Howard A.  
; APPLICANT: Khudiyakov, Yuri E.  
; TITLE OF INVENTION: Antigenically Reactive Regions of the  
; TITLE OF INVENTION: Hepatitis A Virus Polypeptide  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick Stockton LLP  
; STREET: 3424 Peachtree Road, N.E.  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30326  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/171,432A  
; FILING DATE: 23-NOV-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Greene, Jamie L.  
; REGISTRATION NUMBER: 32,457  
; REFERENCE/DOCKET NUMBER: 03063-0231US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 949-2400  
; TELEFAX: (404) 949-2499  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single

```

; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..20
; OTHER INFORMATION: /label= YK-1315
US-09-171-432a-39

Query Match      100.0%; Score 107; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVQKQR 20
Db 1 SHIECRPKYKELRLEVQKQR 20
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |

RESULT 2
US-10-272-459-45
; Sequence 45, Application US/10272459
; Publication No. US20030124517A1
; GENERAL INFORMATION:
; APPLICANT: PICHUANTES, Sergio
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT
; FILE REFERENCE: PPI7955.002 / 2301-17955
; CURRENT APPLICATION NUMBER: US/10/272,459
; CURRENT FILING DATE: 2002-10-15
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 45
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein
US-10-272-459-45

Query Match      100.0%; Score 107; DB 14; Length 352;
Best Local Similarity 100.0%; Pred. No. 4.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVQKQR 20
Db 164 SHIECRPKYKELRLEVQKQR 183
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |

RESULT 3
US-10-272-459-40
; Sequence 40, Application US/10272459
; Publication No. US20030124517A1
; GENERAL INFORMATION:
; APPLICANT: PICHUANTES, Sergio
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT
; FILE REFERENCE: PPI7955.002 / 2301-17955
; CURRENT APPLICATION NUMBER: US/10/272,459
; CURRENT FILING DATE: 2002-10-15
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 40
; LENGTH: 836
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein
US-10-272-459-40

Query Match      100.0%; Score 107; DB 9; Length 2227;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVQKQR 20
Db 1 SHIECRPKYKELRLEVQKQR 20
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |

RESULT 4
US-10-272-459-41
; Sequence 41, Application US/10272459
; Publication No. US20030124517A1
; GENERAL INFORMATION:
; APPLICANT: PICHUANTES, Sergio
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT
; FILE REFERENCE: PPI7955.002 / 2301-17955
; CURRENT APPLICATION NUMBER: US/10/272,459
; CURRENT FILING DATE: 2002-10-15
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 41
; LENGTH: 980
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein
US-10-272-459-41

Query Match      100.0%; Score 107; DB 14; Length 980;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVQKQR 20
Db 792 SHIECRPKYKELRLEVQKQR 811
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |

RESULT 5
US-09-929-955-12
; Sequence 12, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis A virus sequence
US-09-929-955-12

Query Match      100.0%; Score 107; DB 9; Length 2227;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRPKYKELRLEVQKQR 20
Db 1 SHIECRPKYKELRLEVQKQR 20
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
```

```
Db      792 SHIECRKPYKELRLEVQKOR 811
|||||
RESULT 6
US-10-104-966-12
; Sequence 12, Application US/10104966
; Publication No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Salberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; FILE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis A virus sequence
US-10-104-966-12

Query Match      100.0%; Score 107; DB 13; Length 2227;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SHIECRKPYKELRLEVQKOR 20
|||||
Db      792 SHIECRKPYKELRLEVQKOR 811
|||||

RESULT 7
US-10-135-988-2
; Sequence 2, Application US/10135988
; Publication No. US20020176869A1
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US3
; CURRENT APPLICATION NUMBER: US/10/135,988
; CURRENT FILING DATE: 2002-04-29
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-10-135-988-2

Query Match      100.0%; Score 107; DB 13; Length 2227;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SHIECRKPYKELRLEVQKOR 20
|||||
Db      792 SHIECRKPYKELRLEVQKOR 811
|||||

RESULT 8
US-10-135-988-4
; Sequence 4, Application US/10135988
; Publication No. US20020176869A1
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US3
; CURRENT APPLICATION NUMBER: US/10/135,988
; CURRENT FILING DATE: 2002-04-29
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-10-135-988-4

Query Match      100.0%; Score 107; DB 13; Length 2227;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SHIECRKPYKELRLEVQKOR 20
|||||
Db      792 SHIECRKPYKELRLEVQKOR 811
|||||

RESULT 9
US-10-135-988-6
; Sequence 6, Application US/10135988
; Publication No. US20020176869A1
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US3
; CURRENT APPLICATION NUMBER: US/10/135,988
; CURRENT FILING DATE: 2002-04-29
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-10-135-988-6

Query Match      100.0%; Score 107; DB 13; Length 2227;
Best Local Similarity 100.0%; Pred. No. 3.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SHIECRKPYKELRLEVQKOR 20
|||||
Db      792 SHIECRKPYKELRLEVQKOR 811
|||||

RESULT 10
US-09-171-432A-40
; Sequence 40, Application US/0917432A
; Publication No. US20030187184A1
; GENERAL INFORMATION:
```



```

; Sequence 2, Application US/09728912
; Patent No. US20010036643A1
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: Tumor Necrosis Factor-Stimulated Gene
; FILE REFERENCE: 99-94US
; CURRENT APPLICATION NUMBER: US/09/728,912
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: 60/169,252
; PRIOR FILING DATE: 1999-12-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-728-912-2

Query Match      40.2%; Score 43; DB 9; Length 413;
Best Local Similarity 53.3%; Pred. No. 1.2e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      5 CRKPYKELRLVGVKQ 19
      |||:|:|:|:|
Db      376 CRNPTRELRLVREGVQ 390

RESULT 15
US-09-922-217-1116
; Sequence 1116, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yudi
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1116
; LENGTH: 5405
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1116

Query Match      39.7%; Score 42.5; DB 9; Length 5405;
Best Local Similarity 39.1%; Pred. No. 2.2e+03;
Matches 9; Conservative 5; Mismatches 4; Indels 5; Gaps 1;

QY      2 HIECRK----PYKELRLVGVKQ 19
      |:|:|:|:|:|:|
Db      4824 HVTQEGGAACGPHPCRLEDGVQ 4846

Search completed: March 15, 2004, 13:53:26
Job time : 23.1765 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2004, 13:49:12 ; Search time 29.1765 Seconds  
(without alignments)  
216.283 Million cell updates/sec

Title: US-09-171-432A-39  
Perfect score: 107  
Sequence: 1 SHIECRKPKYKELRLEVQKOR 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp arches:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	53	12 Q98VY2	Q98VY2 hepatitis a
2	107	100.0	53	12 Q98VY3	Q98VY3 hepatitis a
3	107	100.0	53	12 Q98VY4	Q98VY4 hepatitis a
4	107	100.0	53	12 Q98VY0	Q98VY0 hepatitis a
5	107	100.0	53	12 Q98VY5	Q98VY5 hepatitis a
6	107	100.0	53	12 Q98VY6	Q98VY6 hepatitis a
7	107	100.0	53	12 Q98VY1	Q98VY1 hepatitis a
8	107	100.0	53	12 Q98VY7	Q98VY7 hepatitis a
9	107	100.0	55	12 Q99V72	Q99V72 hepatitis a
10	107	100.0	55	12 Q99V74	Q99V74 hepatitis a
11	107	100.0	55	12 Q99V73	Q99V73 hepatitis a
12	107	100.0	55	12 Q99V75	Q99V75 hepatitis a
13	107	100.0	55	12 Q99V76	Q99V76 hepatitis a
14	107	100.0	56	12 Q98VY4	Q98VY4 hepatitis a
15	107	100.0	56	12 Q98VY3	Q98VY3 hepatitis a
16	107	100.0	56	12 Q99V71	Q99V71 hepatitis a

ALIGNMENTS

RESULT 1

Q98VY2 ID Q98VY2 PRELIMINARY; PRT; 53 AA.  
AC Q98VY2  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Uruguay6;  
RA Costa-Mattoli M., Ferre V., Monpocho S., Garcia L., Colina R.,  
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic variability of Hepatitis A Virus in South America reveals  
RT heterogeneity and co-circulation during epidemic outbreaks.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ309232; CAC37078.1; -.  
FT NON\_TER 1  
FT NON\_TER 53  
SQ SEQUENCE 53 AA; 6271 MW; 0846AEF4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;  
Best Local Similarity 100.0%; Pred. No. 3.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPKYKELRLEVQKOR 20

Db 29 SHIECRKPKYKELRLEVQKOR 48

RESULT 2

Q98VY3 ID Q98VY3 PRELIMINARY; PRT; 53 AA.  
AC Q98VY3  
DT 01-JUN-2001 (TREMBlrel. 17, Created)

```
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12092;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Urul16;
RA Costa-Mattioli M., Ferre V., Monpoeho S., Garcia L., Colina R.,
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;
RT "Genetic variability of Hepatitis A Virus in South America reveals
RT heterogeneity and co-circulation during epidemic outbreaks.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ309231; CAC37077.1; -.
FT NON_TER 1 1
FT NON_TER 53 53
SQ SEQUENCE 53 AA; 6271 MW; 0846AEF4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SHIECKPKYKELRLEVGKQR 20
Db 29 SHIECKPKYKELRLEVGKQR 48

RESULT 3
Q98VY4 ID Q98VY4 PRELIMINARY; PRT; 53 AA.
AC Q98VY4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12092;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Chile-J;
RA Costa-Mattioli M., Ferre V., Monpoeho S., Garcia L., Colina R.,
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;
RT "Genetic variability of Hepatitis A Virus in South America reveals
RT heterogeneity and co-circulation during epidemic outbreaks.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ309230; CAC37076.1; -.
FT NON_TER 1 1
FT NON_TER 53 53
SQ SEQUENCE 53 AA; 6271 MW; 0846AEF4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SHIECKPKYKELRLEVGKQR 20
Db 29 SHIECKPKYKELRLEVGKQR 48

RESULT 4
Q98VY0 ID Q98VY0 PRELIMINARY; PRT; 53 AA.
AC Q98VY0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
```

```
OC Hepatovirus.
OC NCBI_TaxID=12092;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Uruguay4;
RA Costa-Mattioli M., Ferre V., Monpoeho S., Garcia L., Colina R.,
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;
RT "Genetic variability of Hepatitis A Virus in South America reveals
RT heterogeneity and co-circulation during epidemic outbreaks.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ309234; CAC37080.1; -.
FT NON_TER 1 1
FT NON_TER 53 53
SQ SEQUENCE 53 AA; 6271 MW; 0846AEF4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SHIECKPKYKELRLEVGKQR 20
Db 29 SHIECKPKYKELRLEVGKQR 48

RESULT 5
Q98VY5 ID Q98VY5 PRELIMINARY; PRT; 53 AA.
AC Q98VY5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12092;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Urul13;
RA Costa-Mattioli M., Ferre V., Monpoeho S., Garcia L., Colina R.,
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;
RT "Genetic variability of hepatitis A virus in South America reveals
RT heterogeneity and co-circulation during epidemic outbreaks.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ309229; CAC37075.1; -.
FT NON_TER 1 1
FT NON_TER 53 53
SQ SEQUENCE 53 AA; 6271 MW; 0846AEF4BC397432 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 53;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SHIECKPKYKELRLEVGKQR 20
Db 29 SHIECKPKYKELRLEVGKQR 48

RESULT 6
Q98VY6 ID Q98VY6 PRELIMINARY; PRT; 53 AA.
AC Q98VY6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Urul17;
```



RA Costa-Mattioli M., Ferre V., Monpoeho S., Garcia L., Colina R.,  
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic variability of hepatitis A virus in South America reveals  
RT heterogeneity and co-circulation during epidemic outbreaks";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ309228; CAC37074.1; -.  
FT NON TER 1 1  
SQ SEQUENCE 53 AA; 6271 MW; 0846AEF4BC397432 CRC64;  
Query Match 100.0%; Score 107; DB 12; Length 53;  
Best Local Similarity 100.0%; Pred. No. 3.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECRKPKYKELRLVGVGKOR 20  
DB 29 SHIECRKPKYKELRLVGVGKOR 48  
RESULT 7  
Q98VY1 PRELIMINARY; PRT; 53 AA.  
AC Q98VY1; (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Chile1;  
RA Costa-Mattioli M., Ferre V., Monpoeho S., Garcia L., Colina R.,  
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic variability of hepatitis A virus in South America reveals  
RT heterogeneity and co-circulation during epidemic outbreaks";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ309233; CAC37079.1; -.  
FT NON TER 1 1  
SQ SEQUENCE 53 AA; 6271 MW; 0846AEF4BC397432 CRC64;  
Query Match 100.0%; Score 107; DB 12; Length 53;  
Best Local Similarity 100.0%; Pred. No. 3.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECRKPKYKELRLVGVGKOR 20  
DB 29 SHIECRKPKYKELRLVGVGKOR 48  
RESULT 8  
Q98VY7 PRELIMINARY; PRT; 53 AA.  
AC Q98VY7; (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Urul;  
RA Costa-Mattioli M., Ferre V., Monpoeho S., Garcia L., Colina R.,  
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic variability of hepatitis A virus in South America reveals  
RT heterogeneity and co-circulation during epidemic outbreaks";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ309227; CAC37073.1; -.  
FT NON TER 1 1  
SQ SEQUENCE 53 AA; 6271 MW; 0846AEF4BC397432 CRC64;  
Query Match 100.0%; Score 107; DB 12; Length 53;  
Best Local Similarity 100.0%; Pred. No. 3.8e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECRKPKYKELRLVGVGKOR 20  
DB 29 SHIECRKPKYKELRLVGVGKOR 48  
RESULT 9  
Q99T2 PRELIMINARY; PRT; 55 AA.  
AC Q99T2; (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Urul;  
RA Costa-Mattioli M., Ferre V., Monpoeho S., Garcia L., Colina R.,  
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic variability of hepatitis C virus reveals heterogeneity and  
RT co-circulation during epidemic outbreaks";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ306386; CAC29235.1; -.  
DR FIR; PQ0427; PQ0428.  
DR FIR; PQ0428; PQ0428.  
DR FIR; PQ0430; PQ0430.  
DR FIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target\_S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON TER 1 1  
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;  
Query Match 100.0%; Score 107; DB 12; Length 55;  
Best Local Similarity 100.0%; Pred. No. 3.9e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SHIECRKPKYKELRLVGVGKOR 20  
DB 28 SHIECRKPKYKELRLVGVGKOR 47  
RESULT 10  
Q999T4 PRELIMINARY; PRT; 55 AA.  
AC Q999T4; (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Chile-3;  
RA Costa-Mattioli M., Ferre V., Monpoeho S., Garcia L., Colina R.,  
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic variability of hepatitis C virus reveals heterogeneity and  
RT co-circulation during epidemic outbreaks";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ306384; CAC29233.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON\_TER 1 55  
FT NON\_TER 55  
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;  
  
Query Match 100.0%; Score 107; DB 12; Length 55;  
Best Local Similarity 100.0%; Pred. No. 3.9e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SHIECRKPKYKELRLEVGKOR 20  
Db 28 SHIECRKPKYKELRLEVGKOR 47  
  
RESULT 11  
Q999T3 ID Q999T3 PRELIMINARY; PRT; 55 AA.  
AC Q999T3; 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Chile-9;  
RA Costa-Mattioli M., Ferre V., Monphoelo S., Garcia L., Colina R.,  
Billaudel S., Vega I., Perez-Barcoff R., Cristina J.,  
"Genetic variability of Hepatitis C virus reveals heterogeneity and  
co-circulation during epidemic outbreaks.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ306385; CAC29234.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON\_TER 1 55  
FT NON\_TER 55  
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;  
  
Query Match 100.0%; Score 107; DB 12; Length 55;  
Best Local Similarity 100.0%; Pred. No. 3.9e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SHIECRKPKYKELRLEVGKOR 20  
Db 28 SHIECRKPKYKELRLEVGKOR 47  
  
RESULT 12  
Q999U8 ID Q999U8 PRELIMINARY; PRT; 55 AA.  
AC Q999U8; 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Manuveelo;  
RA Treambonlers A., Jantaradsamee P., Poovorawan Y.,  
"Molecular characterization of Hepatitis A virus infection of an  
outbreak in the southern part of Thailand.";  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF507065; AAM33425.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON\_TER 1 55  
FT NON\_TER 55  
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;  
  
Query Match 100.0%; Score 107; DB 12; Length 55;  
Best Local Similarity 100.0%; Pred. No. 3.9e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SHIECRKPKYKELRLEVGKOR 20  
Db 28 SHIECRKPKYKELRLEVGKOR 47  
  
RESULT 13  
Q8JYF4 ID Q8JYF4 PRELIMINARY; PRT; 56 AA.  
AC Q8JYF4; 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Manuveelo;  
RA Treambonlers A., Jantaradsamee P., Poovorawan Y.,  
"Molecular characterization of Hepatitis A virus infection of an  
outbreak in the southern part of Thailand.";  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF507065; AAM33425.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;  
  
Query Match 100.0%; Score 107; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 4e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SHIECRKPKYKELRLEVGKOR 20  
Db 29 SHIECRKPKYKELRLEVGKOR 48  
  
RESULT 14  
Q67822 ID Q67822 PRELIMINARY; PRT; 56 AA.  
AC Q67822; 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ306384; CAC29233.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON\_TER 1 55  
FT NON\_TER 55  
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;  
  
Query Match 100.0%; Score 107; DB 12; Length 55;  
Best Local Similarity 100.0%; Pred. No. 3.9e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SHIECRKPKYKELRLEVGKOR 20  
Db 28 SHIECRKPKYKELRLEVGKOR 47  
  
RESULT 13  
Q8JYF4 ID Q8JYF4 PRELIMINARY; PRT; 56 AA.  
AC Q8JYF4; 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Manuveelo;  
RA Treambonlers A., Jantaradsamee P., Poovorawan Y.,  
"Molecular characterization of Hepatitis A virus infection of an  
outbreak in the southern part of Thailand.";  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF507065; AAM33425.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;  
  
Query Match 100.0%; Score 107; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 4e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 SHIECRKPKYKELRLEVGKOR 20  
Db 29 SHIECRKPKYKELRLEVGKOR 48  
  
RESULT 14  
Q67822 ID Q67822 PRELIMINARY; PRT; 56 AA.  
AC Q67822; 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]\_TaxID=12092;  
RP SEQUENCE FROM N.A.  
RC STRAIN=ISR-70;  
RA Karetnyi Y.V., Shulman L.M., Manor J., Leitner L., Shehab S.,  
RA Mendelson E.;  
RT "Genetic classification of hepatitis A virus strains isolated in  
RT Israel, based on their VP1/2A nucleotide sequence";  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z77247; CAB01040.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT CHAIN 1 >28 CAPSID PROTEIN VP1.  
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.  
FT NON\_TER 56 56  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 107; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 4e-10; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPKYKELRLEVGKOR 20  
DB 29 SHIECRKPKYKELRLEVGKOR 48

RESULT 15  
O39872 PRELIMINARY; PRT; 56 AA.  
AC O39872;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]\_TaxID=12092;  
RP SEQUENCE FROM N.A.  
RC STRAIN=VDM;  
RX MEDLINE=97247817; PubMed=9039340;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus:  
RT 1982-1996";  
RL J. Med. Virol. 51:273-279(1997).  
RL EMBL; U68697; AAB53593.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT CHAIN 1 >56  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 107; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 4e-10; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SHIECRKPKYKELRLEVGKOR 20  
DB 29 SHIECRKPKYKELRLEVGKOR 48

GenCore version 5.1.6 Copyright (c) 1993 - 2004 Compugen Ltd.									
OM protein - protein search, using sw model									
Run on: May 11, 2004, 13:23:11 ; Search time 47 Seconds (without alignments) 120.233 Million cell updates/sec									
Title: US-09-171-432A-40									
Perfect score: 102									
Sequence: 1 PYKELRLEVQKRLKYAEE 20									
Scoring table: BLOSUM62									
Gapop 10.0 , Gapext 0.5									
Searched: 1586107 seqs, 282547505 residues									
Total number of hits satisfying chosen parameters: 1586107									
Minimum DB seq length: 0									
Maximum DB seq length: 2000000000									
Post-processing: Minimum Match 0%									
Maximum Match 100%									
Listing first 100 summaries									
Database : A_Geneseq_29Jan04.*									
1: Geneseqp1980s.*									
2: Geneseqp1990s.*									
3: Geneseqp2000s.*									
4: Geneseqp2001s.*									
5: Geneseqp2002s.*									
6: Geneseqp2003as.*									
7: Geneseqp2003bs.*									
8: Geneseqp2004s.*									
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.									
SUMMARIES									
Result No.	Score	Query Match	Length	DB ID	Description				
1	102	100.0	20	2	AAW42923 Immunogen				
2	102	100.0	21	4	AAAB69440 Synthetic				
3	102	100.0	366	1	AAP50230 Sequence				
4	102	100.0	854	1	AAP50287 Hepatitis				
5	102	100.0	993	1	AAP50116 Sequence				
6	102	100.0	993	1	AAP50231 Sequence				
7	102	100.0	1077	2	AAW95559 A partial				
8	102	100.0	1091	2	AAAR32426 Translate				
9	102	100.0	2027	2	AAAR05697 Attenuate				
10	102	100.0	2227	2	AAW34074 Hepatitis				
11	102	100.0	2227	3	AAAB18609 Amino aci				
12	102	100.0	2227	3	AAAB18607 Amino aci				
13	102	100.0	2227	3	AAAB18608 Amino aci				
14	102	100.0	2227	5	AAAE19899 Hepatitis				
15	102	100.0	2227	5	AAAG31729 Attenuate				
16	102	100.0	2227	5	AAAG31727 Wild-type				
17	102	100.0	2227	5	AAAG31728 Hepatitis				
18	102	100.0	2227	6	AAU08640 Attenuat				
19	102	100.0	2227	6	AAU08641 Attenuat				
20	102	100.0	2227	6	AAU08639 Wild type				
21	102	100.0	2227	7	AAW00350 Hepatitis				
22	99	97.1	2227	1	AAPE0066 Sequence				
23	94	92.2	839	2	AAAR15629 Capsid re				
24	67	65.7	20	2	AAW42922 Immunogen				
25	67	65.7	21	4	AAAB69439 Synthetic				

99 41 40.2 859 6 ABU38058 Protein e  
100 40.5 39.7 360 4 ABG00270 Novel hum

## ALIGNMENTS

```
RESULT 1
AAW42923
ID AAW42923 standard; peptide; 20 AA.
XX
XX
AC AAW42923;
XX
XX
DT 28-APR-1998 (first entry)
XX
XX
DE Immungenic Hepatitis A virus peptide YK-1316.
XX
XX
KW Immungenic peptide; immunogenic epitope; P2A protein; immune response;
KW antibody.
XX
XX
OS Synthetic.
OS Hepatitis A virus.
XX
XX
FN WO9740147-A1.
XX
XX
PD 30-OCT-1997.
XX
XX
PF 18-APR-1997; 97WO-US006891.
XX
XX
PR 19-APR-1996; 96US-0015644P.
XX
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX
PI Fields HA, Khudyakov YE;
XX
XX
DR WPI; 1997-535831/49.
XX
XX
PT Immungenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune
PT response to HAV in a mammal or to detect the presence of antibodies
PT against HAV in a mammal.
XX
XX
PS Claim 18; Page 112; 140pp; English.
XX
XX
CC Peptides AAW42922-30 are immunogenic peptides corresponding to
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are
CC substantially similar to a portion of the amino acid sequence of the P2A
CC protein of HAV corresponding to amino acids 792-980. The present peptide
CC is derived from amino acids 799-818, and has a reactivity of 41.7% with
CC acute sera. Compositions containing the peptides can be used to induce an
CC immune response to HAV in a mammal. The peptides can also be used to
CC detect the presence of antibodies against HAV in mammalian serum. The
CC peptides can also be used to make an antibody against HAV by
CC administering the peptide to a mammal
XX
XX
SQ Sequence 20 AA;
Query Match 100.0%; Score 102; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PYKELRLEVGKQRLKYAQEE 20
DB 1 PYKELRLEVGKQRLKYAQEE 20
| | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | |
RESULT 2
AAB69440
ID AAB69440 standard; peptide; 21 AA.
XX
XX
AC AAB69440;
XX
XX
DT 20-APR-2001 (first entry)
XX
XX
```

```
DE Synthetic HAV P2A peptide, SEQ ID NO: 40.
XX
XX
KW Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;
KW antigen; major structural capsid polypeptide; HAV antibody detection.
XX
XX
OS Hepatitis A virus.
OS Synthetic.
XX
XX
FN WO200105824-A2.
XX
XX
PD 25-JAN-2001.
XX
XX
PF 14-JUL-2000; 2000WO-US019267.
XX
XX
PR 15-JUL-1999; 99US-0144412P.
XX
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX
PI Fields HA, Khudyakov YE;
XX
XX
DR WPI; 2001-112681/12.
XX
XX
PT Synthetic peptides used as antigen sources for enzyme immunoassays
PT detecting anti-hepatitis A virus and as vaccines.
XX
XX
PS Claim 13; Page 93; 130pp; English.
XX
XX
CC The present sequence is one of a number of synthetic peptides which are
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides
CC comprise antigenic epitopes of the major structural capsid polypeptides
CC or non-structural polypeptides of HAV with one or more glutamine
CC molecules at the carboxy end of the peptide. The peptides are used to
CC detect the presence of antibodies against HAV in mammalian serum, to
CC detect the presence of HAV in a human or animal through the binding of
CC the peptide to an antibody, to detect acute phase infection by detecting
CC IGM antibodies in mammalian serum and detecting convalescence in a
CC mammal. The peptides are used to detect or quantify HAV antibodies in
CC samples in clinical or research-based assays using immunoblotting,
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,
CC tracking of radioactive or bioluminescent markers, chromatography or
CC electrophoresis. The peptides are used to induce an immune response to
CC HAV when administered to a human or animal. Glutamine at the carboxy end
CC of the peptides enhances the IGM antibody reactivity
XX
XX
SQ Sequence 21 AA;
Query Match 100.0%; Score 102; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PYKELRLEVGKQRLKYAQEE 20
DB 1 PYKELRLEVGKQRLKYAQEE 20
| | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | |
RESULT 3
AAP50230
ID AAP50230 standard; protein; 366 AA.
XX
XX
AC AAP50230;
XX
XX
DT 28-NOV-1991 (first entry)
XX
XX
DE Sequence of hepatitis A virus (HAV) surface protein (VP-1).
XX
XX
KW Hepatitis A virus vaccine; immunisation; monoclonal antibody;
KW diagnostic assay.
XX
XX
OS Hepatitis A virus.
XX
XX
FN EPI38704-A.
XX
XX
PD 24-APR-1985.
```

XX 09-OCT-1984; 84EP-00402025.  
XX 14-OCT-1983; 83US-00541836.  
XX 02-MAR-1984; 84US-00585942.  
XX (MERI ) MERCK & CO INC.  
XX Hughes JV, Scolnick EM, Tomassini JE;  
XX WPI; 1985-100818/17.  
XX N-PSDB; AAN50274.  
XX New hepatitis A virus surface protein - useful for binding to  
XX neutralising antibodies to the virus.  
XX Claim 21; Page 46-48; 49pp; English.  
XX VP1 is isolated by solubilisation of the intact virus in an aq. anionic  
XX surfactant and a reducing agent. The viral proteins are sepd. and the  
XX protein of molecular wt. 33000 daltons is sepd  
XX Sequence 366' AA;  
Query Match 100.0%; Score 102; DB 1; Length 366;  
Best Local Similarity 100.0%; Pred. No. 3.9e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVKGKRLKYAOEE 20  
Db 308 PYKELRLEVKGKRLKYAOEE 327  
RESULT 4  
AAP50287  
ID AAP50287 standard; protein; 854 AA.  
XX AAP50287;  
XX 25-MAR-2003 (revised)  
XX 30-NOV-1991 (first entry)  
XX Hepatitis A virus (HAV) peptide corresponding to the capsid protein  
XX region of poliovirus RNA.  
XX Hepatitis A virus assay; antigen; antibody.  
XX Hepatitis A virus.  
XX WO8501517-A.  
XX 11-APR-1985.  
XX 27-SEP-1984; 84WO-US001552.  
XX 30-SEP-1983; 83US-00537911.  
XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX Ticehurst JR, Baltimore D, Feinstein SM, Purcell RH;  
XX Racaniello VR;  
XX WPI; 1985-098846/16.  
XX N-PSDB; AAN50330.  
XX New hepatitis A virus CDNA - useful in assays for the virus and for  
XX prodn. of the viral antigen and antibodies to it.  
XX Example; Fig 7; 60pp; English.  
XX The inventors claim HAV cDNA and a method for producing it, whereby large  
XX amts. can be obtd. economically. The cDNA is useful in the assay for  
XX detection of HAV quickly and easily and with high sensitivity and

CC specificity. The HAV cDNA is also used in the prodn. of HAV antigen or  
CC antibodies to it. The antibodies may be monoclonal. (Updated on 25-MAR-  
CC 2003 to correct PA field.)  
XX Sequence 854 AA;  
Query Match 100.0%; Score 102; DB 1; Length 854;  
Best Local Similarity 100.0%; Pred. No. 9.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVKGKRLKYAOEE 20  
Db 799 PYKELRLEVKGKRLKYAOEE 818  
RESULT 5  
AAP50116  
ID AAP50116 standard; protein; 993 AA.  
XX AAP50116;  
XX 24-OCT-2003 (revised)  
XX 25-MAR-2003 (revised)  
XX 30-SEP-1991 (first entry)  
XX Sequence of Hepatitis A virus (HAV) immunogenic peptides VP-1, VP-2, VP-3  
XX and VP-4.  
XX Antigenic protein; immunogen; vaccine.  
XX Hepatitis A virus; (strain CR326).  
XX EP154587-A.  
XX 11-SEP-1985.  
XX 27-FEB-1985; 85EP-00400369.  
XX 02-MAR-1984; 84US-00585818.  
XX (MERI ) MERCK & CO INC.  
XX Linemeyer DL, Menke JG, Rueben RG, Mitra SW;  
XX WPI; 1985-224964/37.  
XX N-PSDB; AAN50139.  
XX New nucleotide sequences coding for hepatitis A virus antigens - useful  
XX for eliciting normal immune response and in vaccines for protecting  
XX against the virus.  
XX Example; Page 11-17; 32pp; English.  
XX Within the sequence in AAN50139 is encoded the information necessary to  
XX make the antigenic proteins of HAV. The sequences encoding for the  
XX structural proteins begin at base 403. The key sub-unit sequences within  
XX VP-1, designated Sequences I, II, III, IV, and V, start, respectively at  
XX 1882, 1963, 1999, 2146, 2347. Other nucleotide sequences which are  
XX valuable as encoding antigenic proteins are the sequences from base 1749  
XX to base 2722; from base 1487 to base 2980 and from base 1644 to base  
XX 2722. The sequence from base 1749 to base 2722 is esp. valuable as a  
XX vector for producing antigen protein. Sequences II-V are claimed. X in  
XX AAP50116 denotes the translation of a stop codon. (Updated on 25-MAR-2003  
XX to correct PI field.) (Updated on 24-OCT-2003 to standardise OS field)  
XX Sequence 993 AA;  
Query Match 100.0%; Score 102; DB 1; Length 993;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVKGKRLKYAOEE 20  
|||

Db 935 PYKELRLEVGKQRLKYAQEE 954

RESULT 6  
AAP50231  
ID AAP50231 standard; protein; 993 AA.  
XX AC AAP50231;  
XX 28-NOV-1991 (first entry)  
DT DT  
XX Sequence encoded by partial sequence of hepatitis A virus (HAV),  
DE including surface protein (VP-1).  
XX Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
KW diagnostic assay.  
XX Hepatitis A virus.  
OS  
XX Key Location/Qualifiers  
FH Protein 628..993  
FT /note= "claimed; X denotes translated stop codons and  
FT unspecified triplets"  
XX EPI38704-A.  
FN 24-APR-1985.  
XX 09-OCT-1984; 84EP-00402025.  
XX 14-OCT-1983; 83US-00541836.  
PR 02-MAR-1984; 84US-00585942.  
XX (MERI ) MERCK & CO INC.  
PA  
XX Hughes JV, Scolnick EM, Tomassini JE;  
PI WPI; 1985-100818/17.  
XX N-PSDB; AAN50274.  
XX New hepatitis A virus surface protein - useful for binding to  
PT neutralising antibodies to the virus.  
XX Disclosure; Page 17-23; 49pp; English.  
XX VPI is isolated by solubilisation of the intact virus in an aq. anionic  
CC surfactant and a reducing agent. The viral proteins are sepd. and the  
CC protein of molecular wt. 33000 daltons is sepd  
XX Sequence 993 AA;  
SQ  
Query Match 100.0%; Score 102; DB 1; Length 993;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 PYKELRLEVGKQRLKYAQEE 20  
Db 935 PYKELRLEVGKQRLKYAQEE 954

RESULT 7  
AAW95559  
ID AAW95559 standard; protein; 1077 AA.  
XX AC AAW95559;  
XX 28-APR-1999 (first entry)  
DT A partial hepatitis A virus (HAV) protein.  
XX Hepatitis A virus protein; HAV; P2 region;  
KW cell-culture-adapted HAV strain; infection; accelerated growth.  
XX

OS Hepatitis A virus.  
XX US5849562-A.  
PN 15-DEC-1998.  
PD 06-JUN-1995; 95US-00468926.  
XX 30-SEP-1983; 83US-00537911.  
PR 27-SEP-1984; 84US-00654942.  
XX 06-OCT-1988; 88US-00256135.  
PR 06-NOV-1991; 91US-00788262.  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA Emerson SU, Purcell RH;  
XX WPI; 1999-094412/08.  
DR N-PSDB; AAX01006.  
XX Chimeric hepatitis A virus strains - with P2 region from cell-culture-  
PT adapted strain in wild-type genome.  
XX Disclosure; Fig 7A-L; 36pp; English.  
XX The present sequence represents a partial hepatitis A virus (HAV)  
CC protein. The specification describes a DNA construct consisting of a wild  
CC -type HAV genome in which the P2 region is replaced by the P2 region from  
CC a cell-culture-adapted HAV strain. The construct is used to demonstrate  
CC that mutations in the P2 region of a cell-culture-adapted HAV strain are  
CC sufficient for establishment of infection and accelerated growth in cell  
CC culture  
XX Sequence 1077 AA;  
SQ  
Query Match 100.0%; Score 102; DB 2; Length 1077;  
Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 PYKELRLEVGKQRLKYAQEE 20  
Db 1022 PYKELRLEVGKQRLKYAQEE 1041

RESULT 8  
AAR32426  
ID AAR32426 standard; protein; 1091 AA.  
XX AC AAR32426;  
XX 27-AUG-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 17-DEC-2001 (revised)  
DT 10-JUN-1993 (first entry)  
XX Translated from 5' region of Hepatitis A Virus genomic clone.  
DE HAV HM-175; chronic liver disease; picornavirus.  
XX Hepatitis A virus.  
OS  
XX Key Location/Qualifiers  
FH Region 1..711  
FT /note= "X's correspond to nonsense codons, i.e. this  
FT region is not an ORF"  
FT 238..1091  
FT /label= ORF  
FT /note= "second putative initiation codon at position 240"  
XX USN7786262-N.  
XX 15-DEC-1992.  
PD

PF 06-NOV-1991; 91US-00788262.  
 XX 30-SEP-1983; 88US-00536911.  
 PR 27-SEP-1984; 84US-00654942.  
 PR 06-OCT-1988; 88US-00256135.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PA Ticehurst JR, Baltimore D, Feinstein SM, Purcell RH;  
 XX Racanietello VR, Baroudy BM, Emerson SU;  
 PI WPI; 1993-067429/08.  
 DR N-PSDB; AAQ36934.  
 XX Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn. of  
 PT antigen and antibodies.  
 PT Disclosure; Fig 7; 65pp; English.  
 PS HAV virion RNA was extracted from the livers of marmosets which had been  
 CC inoculated with HAV (the HAV had previously been passaged twice in  
 CC marmosets). The RNA was used to prepare ds cDNA clones by standard  
 CC methods. Clones contg. inserts which hybridised to RNA from HAV-infected  
 CC African Green Monkey Kidney cells were selected for further analysis. A  
 CC 7.4kb restriction map (about 99% of the HAV genome) was constructed from  
 CC 5 overlapping inserts. The sequence of the first 3.3kb (approx.) from the  
 CC 5'-terminus was determined. An amino acid sequence was deduced from the  
 CC entire clone and an open reading frame was identified starting at  
 CC position 238. A comparison of the predicted HAV amino acid sequences with  
 CC the known capsid protein sequences of other picornaviruses (poliovirus,  
 CC foot and mouth disease virus and encephalomyelitis virus) revealed areas  
 CC of local homology. (Note: Revised entry submitted to correct the patent  
 CC number format of US Government-owned NTIS applications to prevent clashes  
 CC with ongoing US granted patent numbers. For further information please  
 CC visit the Derwent web site at [www.derwent.com/dwpi/updates/ntis.us.html](http://www.derwent.com/dwpi/updates/ntis.us.html).)  
 CC (Updated on 25-MAR-2003 to correct PI field.) (Updated on 27-AUG-2003 to  
 CC correct OS field.)  
 XX SQ Sequence 1091 AA;  
 Query Match 100.0%; Score 102; DB 2; Length 1091;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07; Indels 0; Gaps 0;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PYKELRLEVGKQRLKYAQEE 20  
 DB 1036 PYKELRLEVGKQRLKYAQEE 1055  
 RESULT 9  
 ID AAR05697 standard; protein; 2227 AA.  
 XX AAR05697;  
 AC AAR05697;  
 XX 24-OCT-2003 (revised)  
 DT 27-AUG-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 15-AUG-1990 (first entry)  
 XX Attenuated hepatitis A virus.  
 DE Hepatitis A virus; vaccine; attenuated.  
 KW Hepatitis A virus; strain HM-175.  
 OS Hepatitis A virus; protein; 2227 AA.  
 XX AAW34074 standard; protein; 2227 AA.  
 ID AAW34074  
 XX AAW34074;  
 AC AAW34074;  
 XX 17-OCT-2003 (revised)  
 DT 27-APR-1998 (first entry)  
 XX Hepatitis A virus HM-175 protein sequence.  
 DE HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus; infection;  
 KW vaccine.

FT Region 492..791  
 FT /label= VP1 = 1D  
 FT Region 792..980  
 FT /label= 2A  
 FT Region 981..1087  
 FT /label= 2B  
 FT Region 1088..1422  
 FT /label= 2C  
 FT Region 1423..1496  
 FT /label= 3A  
 FT Region 1497..1519  
 FT /label= 3B = VPg  
 FT Region 1520..1738  
 FT /label= 3C  
 FT Region 1739..2227  
 FT /label= 3D  
 XX US4894228-A.  
 PN 16-JAN-1990.  
 XX 12-JUL-1988; 88US-00217824.  
 XX 19-SEP-1984; 84US-00652067.  
 PR 09-SEP-1986; 86US-00905146.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PA Purcell RH, Ticehurst JR, Cohen JL, Emerson SU, Feinstein SM;  
 PI Raemer RJ, Gust ID;  
 XX WPI; 1990-075557/10.  
 DR N-PSDB; AAQ03512.  
 XX Vaccine against hepatitis A virus infection - comprises novel attenuated  
 PT hepatitis A virus strain.  
 XX Claim 1; Fig 1; 18pp; English.  
 PS The attenuated HAV is useful for inducing protective immunity against  
 CC HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by  
 CC several nucleotide changes distributed throughout the genome. It is  
 CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is  
 CC suitable for use as an HAV vaccine. It is noted that not all the changes  
 CC are necessary for attenuation and use as a vaccine. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PI field.)  
 CC (Updated on 27-AUG-2003 to correct OS field.) (Updated on 24-OCT-2003 to  
 CC standardise OS field)  
 XX SQ Sequence 2227 AA;  
 Query Match 100.0%; Score 102; DB 2; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PYKELRLEVGKQRLKYAQEE 20  
 DB 799 PYKELRLEVGKQRLKYAQEE 818  
 RESULT 10  
 ID AAW34074  
 XX AAW34074 standard; protein; 2227 AA.  
 ID AAW34074  
 XX AAW34074;  
 AC AAW34074;  
 XX 17-OCT-2003 (revised)  
 DT 27-APR-1998 (first entry)  
 XX Hepatitis A virus HM-175 protein sequence.  
 DE HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus; infection;  
 KW vaccine.



```

RESULT 11
AAB18609
ID AAB18609 standard; protein; 2227 AA.
XX
XX AAB18609;
XX
XX 15-JAN-2001 (first entry)
XX
XX Amino acid sequence of live attenuated Hepatitis A virus 4380.
DE HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
DE HAV 4380.
XX
XX Hepatitis A virus.
OS
XX US6113912-A.
XX
XX 05-SEP-2000.
XX
XX 07-JUN-1995; 95US-00475886.
XX
XX 18-SEP-1992; 92US-00947338.
XX
XX 17-SEP-1993; 93WO-US008610.
XX
XX 17-APR-1995; 95US-00397232.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;
PI
XX WPI; 2000-586464/55.
XX
XX N-PSDB; AAA75478.
XX
XX Novel live hepatitis A virus adapted to growth in human fibroblast cell
XX line useful as vaccine for protecting humans against hepatitis A virus
XX infection, has modified genome compared to wild type.
XX
XX Disclosure; Col 93-104; 72pp; English.
XX
XX The present sequence is derived from a live attenuated hepatitis A virus
XX (HAV) of the invention, designated HAV 4380. The sequence is produced by
XX modifying wild type HAV strain HM-174. The HAV of the invention are
XX adapted to growth in the human fibroblast-like cell line MRC-5. The HAV
XX is able to propagate in MRC-5 cells and retain appropriate attenuation.
XX It is useful as a live vaccine for prophylaxis of hepatitis A in humans
XX and other primates
XX
XX Sequence 2227 AA;
SQ
Query Match 100.0%; Score 102; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2,7e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
Db 799 PYKELRLEVGKQRLKYAQEE 818
|||||
|||||

RESULT 12
AAB18607
ID AAB18607 standard; protein; 2227 AA.
XX
XX AAB18607;
XX
XX 15-JAN-2001 (first entry)
XX
XX Amino acid sequence of wild type Hepatitis A virus strain HM-175.
DE HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.
XX
XX Hepatitis A virus.
XX
XX US6113912-A.
PN

```

```

XX 05-SEP-2000.
XX 07-JUN-1995; 95US-00475886.
XX 18-SEP-1992; 92US-00947338.
XX 17-SEP-1993; 93WO-US008610.
XX 17-APR-1995; 95US-00397232.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX WPI; 2000-586464/55.
XX N-PSDB; AAA75476.
XX Novel live hepatitis A virus adapted to growth in human fibroblast cell
XX line useful as vaccine for protecting humans against hepatitis A virus
XX infection, has modified genome compared to wild type.
XX Disclosure; Fig 6A-K; 72pp; English.
XX The present sequence is derived from a wild type hepatitis A virus (HAV)
XX strain HM-174. The sequence is modified to produce HAV which are adapted
XX to growth in the human fibroblast-like cell line MRC-5. The HAV is able
XX to propagate in MRC-5 cells and retain appropriate attenuation. It is
XX useful as a live vaccine for prophylaxis of hepatitis A in humans and
XX other primates
XX SQ Sequence 2227 AA;
XX Query Match 100.0%; Score 102; DB 3; Length 2227;
XX Best Local Similarity 100.0%; Pred. No. 2.7e-07;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 PYKELRLEVGKQRLKYAQEE 20
XX Db 799 PYKELRLEVGKQRLKYAQEE 818
XX
XX RESULT 13
XX ID AAB18608
XX ID AAB18608 standard; protein; 2227 AA.
XX AC AAB18608;
XX AC XX
XX DT 15-JAN-2001 (first entry)
XX DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.
XX XX HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;
XX KW P-35 virus.
XX XX Hepatitis A virus.
XX OS Hepatitis A virus.
XX XX US6113912-A.
XX XX 05-SEP-2000.
XX XX 07-JUN-1995; 95US-00475886.
XX XX 18-SEP-1992; 92US-00947338.
XX PR 17-SEP-1993; 93WO-US008610.
XX PR 17-APR-1995; 95US-00397232.
XX XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX XX D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;
XX PI WPI; 2000-586464/55.
XX XX N-PSDB; AAA75477.
XX XX Novel live hepatitis A virus adapted to growth in human fibroblast cell
XX PR

```

DT	29-NOV-2002 (first entry)
DE	Wild-type Hepatitis A virus (HAV) strain HM-175 polypeptide.
XX	
XX	Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast; virucide.
XX	
XX	Hepatitis A virus; strain HM-175.
OS	
XX	US6423318-B1.
PN	
XX	
XX	23-JUL-2002.
PD	
XX	
XX	31-AUG-2000; 2000US-00653499.
PR	
XX	17-SEP-1993; 93WO-US008610.
PR	17-APR-1995; 95US-0039232.
PR	07-JUN-1995; 95US-00475886.
PR	
XX	(USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA	(SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA	
XX	
XX	Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;
PI	
XX	WPI; 2002-680946/73.
DR	N-PSDB; ABS52787.
DR	
XX	
XX	New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth
PT	in MRC-5 cells, useful for preparing a vaccine against HAV infection.
PT	
XX	Disclosure; Fig 6; 71pp; English.
PS	
XX	The invention relates to a polynucleotide which encodes a hepatitis A
XX	virus (HAV) adapted to growth in MRC-5 cells (human fibroblast-like cell
CC	line). The polynucleotide is useful for preparing a vaccine against
CC	hepatitis A virus infection. This sequence represents a hepatitis A virus
CC	strain HM-175 polypeptide. (Updated on 29-AUG-2003 to standardise OS
CC	field)
CC	
XX	
SQ	Sequence 2227 AA;
	Query Match 100.0%; Score 102; DB 5; Length 2227;
	Best Local Similarity 100.0%; Pred. NO. 2.7e-07;
	Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 PYKELRLEVGKQLKYAQEE 20
	799 PYKELRLEVGKQLKYAQEE 818
Db	
	RESULT 17
	ABG31728
ID	ABG31728 standard; protein; 2227 AA.
XX	
AC	ABG31728;
XX	
DT	29-NOV-2002 (first entry)
XX	
DE	Hepatitis A virus mutant strain HM-175/7 (pHAV/7) polypeptide.
XX	
XX	Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast;
KW	virucide; mutant; pHAV/7; mutein.
XX	
OS	Hepatitis A virus; strain HM-175.
OS	Synthetic.
XX	
XX	Key Location/Qualifiers
FT	Misc-difference 764
FT	/note= "Wild-type Glu substituted by Val"
FT	Misc-difference 821
FT	/note= "Wild-type Asn substituted by Ser"
FT	Misc-difference 963
FT	/label= Wild-type Lys substituted by Arg
FT	Misc-difference 1052

FT FT /note= "Wild-type Ala substituted by Val"  
Misc-difference 1062  
FT /note= "Wild-type Gly substituted by Ala"  
Misc-difference 1118  
FT /note= "Wild-type Lys substituted by Met"  
Misc-difference 1151  
FT /note= "Wild-type Glu substituted by Lys"  
Misc-difference 1163  
FT /note= "Wild-type Phe substituted by Ser"  
Misc-difference 1277  
FT /note= "Wild-type Val substituted by Ile"  
Misc-difference 1500  
FT /note= "Wild-type His substituted by Tyr"  
Misc-difference 1805  
FT /note= "Wild-type Asp substituted by Asn"  
Misc-difference 1930  
FT /note= "Wild-type Ser substituted by Thr"  
XX  
PN US6423318-B1.  
XX  
XX 23-JUL-2002.  
XX  
XX 31-AUG-2000; 2000US-00653499.  
XX  
XX 17-SEP-1993; 93WO-US008610.  
PR 17-APR-1995; 95US-00397232.  
PR 07-JUN-1995; 95US-00475886.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA (SWIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
PI  
XX WPI; 2002-680946/73.  
DR N-PSDB; ABS52788.  
XX  
XX New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth  
FT in MRC-5 cells, useful for preparing a vaccine against HAV infection.  
PT  
XX Example 3; Col 67-78; 71pp; English.  
PS  
XX The invention relates to a polynucleotide which encodes a hepatitis A  
CC virus (HAV) adapted to growth in MRC-5 cells (human fibroblast-like cell  
CC line). The polynucleotide is useful for preparing a vaccine against  
CC hepatitis A virus infection. This sequence represents a hepatitis A virus  
CC mutant strain HM-175/7 (pHAV/7) polypeptide  
XX  
SQ Sequence 2227 AA;  
Query Match 100.0%; Score 102; DB 5; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
DB 799 PYKELRLEVGKQRLKYAQEE 818  
XX  
RESULT 18  
ABU08640  
ID ABU08640 standard; protein; 2227 AA.  
XX  
XX ABU08640;  
XX  
XX 23-OCT-2003 (revised)  
DT 03-JUN-2003 (first entry)  
XX  
XX Attenuated (pass35) hepatitis A virus strain HM-175.  
DE  
XX Hepatitis A virus; HAV; virucide; hepatotropic; antiinflammatory;  
KW vaccine; MRC-5 cell; hepatitis infection; ds; attenuated virus.  
XX  
XX Hepatitis A virus; strain HM-175.  
OS

XX US2002176869-A1.  
PN  
XX 28-NOV-2002.  
PD  
XX 29-APR-2002; 2002US-00135988.  
XX  
XX 18-SEP-1992; 92US-00947338.  
PR 17-SEP-1993; 93WO-US008610.  
PR 17-APR-1995; 95US-00397232.  
PR 07-JUN-1995; 95US-00475886.  
PR 31-AUG-2000; 2000US-00653499.  
XX  
XX (FUNK/) FUNKHOUSER A W.  
PA (EMER/) EMERSON S U.  
PA (PURC/) PURCELL R H.  
PA (DHON/) D'HONDT E.  
XX  
XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
PI  
XX WPI; 2003-352605/02.  
DR N-PSDB; ABX93474.  
XX  
XX New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells,  
FT useful in vaccines for protecting primates against hepatitis infection  
PT and disease.  
XX  
XX Example 3; Fig 6; 70pp; English.  
PS  
XX The invention describes a live hepatitis A virus (HAV) adapted to growth  
CC in MRC-5 cells. The HAV and compositions comprising the HAV are useful as  
CC a vaccine for protecting primates against hepatitis infection and  
CC disease. This is the amino acid sequence of an attenuated (pass 35)  
CC human hepatitis A virus strain HM-175. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 2227 AA;  
Query Match 100.0%; Score 102; DB 6; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
DB 799 PYKELRLEVGKQRLKYAQEE 818  
XX  
RESULT 19  
ABU08641  
ID ABU08641 standard; protein; 2227 AA.  
XX  
XX ABU08641;  
XX  
XX 23-OCT-2003 (revised)  
DT 03-JUN-2003 (first entry)  
XX  
XX Attenuated hepatitis A virus (4380) strain HM-175.  
DE  
XX Hepatitis A virus; HAV; virucide; hepatotropic; antiinflammatory;  
KW vaccine; MRC-5 cell; hepatitis infection; ds; attenuated virus.  
XX  
XX Hepatitis A virus; strain HM-175.  
XX  
XX US2002176869-A1.  
PN  
XX 28-NOV-2002.  
PD  
XX 29-APR-2002; 2002US-00135988.  
XX  
XX 18-SEP-1992; 92US-00947338.  
PR 17-SEP-1993; 93WO-US008610.  
PR 17-APR-1995; 95US-00397232.  
PR 07-JUN-1995; 95US-00475886.  
PR

PR 31-AUG-2000; 2000US-00653499.  
XX (FUNK/) FUNKHOUSER A W.  
PA (EMER/) EMERSON S U.  
PA (PURC/) PURCELL R H.  
PA (DHON/) D'HONDT E.  
XX  
XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
PI WPI; 2003-352605/02.  
XX N-PSDB; ABX93475.  
DR  
DR New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells,  
XX PT useful in vaccines for protecting primates against hepatitis infection  
PT and disease.  
XX Disclosure; Page 45-51; 70pp; English.  
PS  
XX The invention describes a live hepatitis A virus (HAV) adapted to growth  
CC in MRC-5 cells. The HAV and compositions comprising the HAV are useful as  
CC a vaccine for protecting primates against hepatitis infection and  
CC disease. This is the amino acid sequence of an attenuated human  
CC hepatitis A virus (4380) strain HM-175. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX  
XX Sequence 2227 AA;  
SQ

Query Match 100.0%; Score 102; DB 6; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
DB 799 PYKELRLEVGKQRLKYAQEE 818  
|||||

RESULT 20  
ABU08639  
ID ABU08639 standard; protein; 2227 AA.  
XX  
XX ABU08639;  
AC  
XX 23-OCT-2003 (revised)  
DT 03-JUN-2003 (first entry)  
XX  
XX Wild type human hepatitis A virus strain HM-175.  
DE  
XX Hepatitis A virus; HAV; virucide; hepatotropic; antiinflammatory;  
XX vaccine; MRC-5 cell; hepatitis infection.  
XX  
XX Hepatitis A virus; strain HM-175.  
OS  
XX US2002176869-A1.  
XX  
XX 28-NOV-2002.  
PD

XX 29-APR-2002; 2002US-00135988.  
FF  
XX 18-SEP-1992; 92US-00947338.  
PR 17-SEP-1993; 93WO-US008610.  
XX 17-APR-1995; 95US-00397232.  
PR 07-JUN-1995; 95US-00475886.  
PR 31-AUG-2000; 2000US-00653499.  
XX  
XX (FUNK/) FUNKHOUSER A W.  
PA (EMER/) EMERSON S U.  
PA (PURC/) PURCELL R H.  
PA (DHON/) D'HONDT E.  
XX  
XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
PI WPI; 2003-352605/02.  
DR N-PSDB; ABX93473.  
DR

XX New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells,  
PT useful in vaccines for protecting primates against hepatitis infection  
PT and disease.  
XX  
XX Disclosure; Fig 6; 70pp; English.  
XX  
XX The invention describes a live hepatitis A virus (HAV) adapted to growth  
CC in MRC-5 cells. The HAV and compositions comprising the HAV are useful as  
CC a vaccine for protecting primates against hepatitis infection and  
CC disease. This is the amino acid sequence of wild type human hepatitis A  
CC virus strain HM-175. (Updated on 23-OCT-2003 to standardise OS field)  
XX  
XX Sequence 2227 AA;  
SQ

Query Match 100.0%; Score 102; DB 6; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
DB 799 PYKELRLEVGKQRLKYAQEE 818  
|||||

RESULT 21  
ABW00350  
ID ABW00350 standard; protein; 2227 AA.  
XX  
XX ABW00350;  
AC  
XX 15-JAN-2004 (first entry)  
DT  
XX Hepatitis A virus protein.  
DE  
XX Ribavirin; vaccine; immune response; infection; therapy; immunostimulant;  
XX virucide.  
KW  
XX Hepatitis A virus.  
OS  
XX US2002136740-A1.  
PN  
XX 26-SEP-2002.  
PD  
XX 15-AUG-2001; 2001US-00929955.  
PF  
XX 17-AUG-2000; 2000US-0225767P.  
PR 29-AUG-2000; 2000US-0229175P.  
PR  
XX (SALL/) SALLBERG M.  
PA (HULT/) HULTGREN C.  
XX  
XX Sallberg M, Hultgren C;  
PI  
XX WPI; 2003-764978/72.  
DR N-PSDB; AAD60867.  
DR  
XX Vaccine compositions for treating and preventing disease, preferably  
PT hepatitis C virus infection, comprises ribavirin and antigen that has  
PT epitope present in hepatitis C virus.  
XX  
XX Claim 11; Page 45-51; Opp; English.  
PS  
XX The invention relates to a composition comprising ribavirin and an  
CC antigen, where the antigen is derived from a hepatitis virus. The vaccine  
CC is useful in enhancing the immune response to a hepatitis C antigen where  
CC the composition is delivered to an animal identified as requiring an  
CC enhanced immune response. The vaccine is useful in the treatment and  
CC prevention of hepatitis C infection. The present sequence is Hepatitis A  
CC virus protein  
XX  
XX Sequence 2227 AA;  
SQ

Query Match 100.0%; Score 102; DB 7; Length 2227;

Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLVGVKQRLKYAQEE 20  
| | | | | | | | | | | | | | | | | | | | | |  
DB 799 PYKELRLVGVKQRLKYAQEE 818

RESULT 22  
AAF60066  
ID AAF60066 standard; protein; 2227 AA.  
XX AC AAF60066;  
XX 25-MAR-2003 (revised)  
XX 26-JUN-1991 (first entry)  
XX DE Sequence of viral I434 polypeptide encoded by the complete nucleotide  
XX sequence of the HAV genome.  
XX DE  
XX Diagnosis; vaccine; passive immunotherapy.  
XX KW  
XX OS Hepatitis A virus.  
XX XX  
XX Key Location/Qualifiers  
XX 1. 245  
XX Region /label= P1.1A  
XX FT 246..491  
XX FT Region /label= 1B  
XX FT 492..836  
XX FT Region /label= 1C  
XX FT 837..980  
XX FT Region /label= P2.2A  
XX FT 981..1076  
XX FT Region /label= 2B  
XX FT 1077..1422  
XX FT Region /label= 2C  
XX FT 1423..1484  
XX FT Region /label= P3.3A  
XX FT 1485..1507  
XX FT Region /label= 3B  
XX FT 1508..1678  
XX FT Region /label= 3C  
XX FT 1679..2227  
XX FT Region /label= 3D  
XX PN EP199480-A.  
XX XX  
XX 29-OCT-1986.  
XX XX  
XX 03-APR-1986; 86EP-00302465.  
XX PF  
XX 03-APR-1985; 85US-00719329.  
XX PR  
XX (CHIR ) CHIRON CORP.  
XX XX  
XX PI Dina D, Potter SJ, Vannest GA, Caput D;  
XX XX  
XX WPI; 1986-286213/44.  
XX DR N-PSDB; AAN60080.  
XX DR  
XX Hepatitis A virus nucleotide sequence and polypeptide - and use in prodn.  
XX of vaccines and diagnostic probes.  
XX Claim 5; Fig 1; 18pp; English.  
XX XX  
XX AAN60080 and oligonucleotide fragments are useful in detection of  
XX hepatitis A virus; transformed hosts may be used for expression of  
XX polypeptides and fragments useful in vaccines without risk of infection  
XX by the virus or in prodn. of particles which are capable of inducing  
XX immunocompetent B cells for passive immunotherapy. Pref. epitope is  
XX derived from AAS 445-657 or 792-848 of the HAV polypeptide sequence  
XX (AAF60066). (Updated on 25-MAR-2003 to correct PA field.)  
XX CC

CC cyto-HAV. The other two cleavage sites are the same. Two residues have  
 CC been identified as part of the immuno- dominant region (see feature  
 CC table) and are different to those in the same position in human HAV. The  
 CC protein and peptides derived from it can be used in the prepn. of  
 CC vaccines for the prevention of HAV infection. See also AAR15056. (Note:  
 CC Revised entry submitted to correct the patent number format of US  
 CC Government-owned NTIS applications to prevent clashes with ongoing US  
 CC granted patent numbers. For further information please visit the Derwent  
 CC web site at [www.derwent.com/dwpi/updates/ntis.us.html](http://www.derwent.com/dwpi/updates/ntis.us.html).) (Updated on 25-  
 CC MAR-2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS  
 CC field.) (Updated on 24-OCT-2003 to standardise OS field)

XX Sequence 839 AA;

Query Match 92.2%; Score 94; DB 2; Length 839;  
 Best Local Similarity 90.0%; Pred. NO. 1.9e-06;  
 Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PYKELRLVGVKQRLKVAQEE 20  
 |||||  
 Db 798 PYKELRLVGVKQRFKVAEE 817

RESULT 24  
 AAW42922  
 ID AAW42922 standard; peptide; 20 AA.

XX AC AAW42922;

DT 28-APR-1998 (first entry)

XX Immunogenic Hepatitis A virus peptide YK-1315.

XX Immunogenic peptide; immunogenic epitope; P2A protein; immune response;  
 antibody.

OS Synthetic.

OS Hepatitis A virus.

XX WO9740147-A1.

XX 30-OCT-1997.

XX 18-APR-1997; 97WO-US006891.

XX 19-APR-1996; 96US-0015644P.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

XX WPI; 1997-535831/49.

XX Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
 PT response to HAV in a mammal or to detect the presence of antibodies  
 PT against HAV in a mammal.

XX Claim 18; Page 112; 140pp; English.

XX Peptides AAW42922-30 are immunogenic peptides corresponding to  
 CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
 CC substantially similar to a portion of the amino acid sequence of the P2A  
 CC protein of HAV corresponding to amino acids 792-980. The present peptide  
 CC is derived from amino acids 792-811, and has a reactivity of 54.2% with  
 CC acute sera. Compositions containing the peptides can be used to induce an  
 CC immune response to HAV in a mammal. The peptides can also be used to  
 CC detect the presence of antibodies against HAV in mammalian serum. The  
 CC peptides can also be used to make an antibody against HAV by  
 CC administering the peptide to a mammal

XX Sequence 20 AA;

Query Match 65.7%; Score 67; DB 2; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.00086;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLVGVKQRL 13

|||||

Db 8 PYKELRLVGVKQRL 20

RESULT 25

AAAB69439

ID AAB69439 standard; peptide; 21 AA.

XX AC AAB69439;

XX 20-APR-2001 (first entry)

XX Synthetic HAV P2A peptide, SEQ ID NO: 39.

XX Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
 KW antigen; major structural capsid polypeptide; HAV antibody detection.

XX Hepatitis A virus.

OS Synthetic.

XX WO200105824-A2.

XX 25-JAN-2001.

XX 14-JUL-2000; 2000WO-US019267.

XX 15-JUL-1999; 99US-0144412P.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

XX WPI; 2001-112681/12.

XX Synthetic peptides used as antigen sources for enzyme immunoassays  
 PT detecting anti-hepatitis A virus and as vaccines.

XX Claim 13; Page 93; 130pp; English.

XX The present sequence is one of a number of synthetic peptides which are  
 CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
 CC comprise antigenic epitopes of the major structural capsid polypeptides  
 CC or non-structural polypeptides of HAV with one or more glutamine  
 CC molecules at the carboxy end of the peptide. The peptides are used to  
 CC detect the presence of antibodies against HAV in mammalian serum, to  
 CC detect the presence of HAV in a human or animal through the binding of  
 CC the peptide to an antibody, to detect acute phase infection by detecting  
 CC IGM antibodies in mammalian serum and detecting convalescence in a  
 CC mammal. The peptides are used to detect or quantify HAV antibodies in  
 CC samples in clinical or research-based assays using immunoblotting,  
 CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
 CC tracking of radioactive or bioluminescent markers, chromatography or  
 CC electrophoresis. The peptides are used to induce an immune response to  
 CC HAV when administered to a human or animal. Glutamine at the carboxy end  
 CC of the peptides enhances the IGM antibody reactivity

XX Sequence 21 AA;

Query Match 65.7%; Score 67; DB 4; Length 21;

Best Local Similarity 100.0%; Pred. No. 0.00091;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLVGVKQRL 13

|||||

Db 8 PYKELRLVGVKQRL 20

RESULT 26

ADA33461

ADA33461 standard; protein; 249 AA.  
 ADA33461;  
 20-NOV-2003 (first entry)  
 Acinetobacter baumannii protein #622.  
 Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;  
 plant biocontrol agent.  
 Acinetobacter baumannii.  
 US6562958-B1.  
 13-MAY-2003.  
 04-JUN-1999; 95US-00328352.  
 09-JUN-1998; 98US-0088701P.  
 (GENO-) GENOME THERAPEUTICS CORP.  
 Breton G, Bush D;  
 WPI; 2003-576092/54.  
 N-PSDB; ADA29335.  
 New Acinetobacter baumannii proteins and nucleic acids, useful as reagents  
 for diagnosing a bacterial disease, as components of antibacterial  
 vaccines, as targets for antibacterial drugs, or as biocontrol agents for  
 plants.  
 Example; SEQ ID NO 4748; 328pp; English.  
 The invention relates to isolated Acinetobacter baumannii nucleic acids.  
 The A. baumannii nucleic acids and polypeptides are useful as reagents  
 for diagnosing a bacterial disease, as components of antibacterial  
 vaccines, as targets for antibacterial drugs, to detect the presence of  
 A. baumannii and other Acinetobacter species in a sample, in screening  
 compounds for the ability to interfere with the A. baumannii life cycle  
 or to inhibit A. baumannii infection, and as biocontrol agents for  
 plants. The present sequence represents the amino acid sequence of an A.  
 baumannii protein.  
 Sequence 249 AA;  
 Query Match 46.1%; Score 47; DB 6; Length 249;  
 Best Local Similarity 56.2%; Pred. No. 22;  
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 PYKLELVGVQRLKY 16  
 DB 5 PMKSLWLDIGNTRUKY 20  
 RESULT 27  
 ABP31097  
 ID ABP31097 standard; protein; 65 AA.  
 AC ABP31097;  
 08-JUL-2002 (first entry)  
 Human ORF70 protein, SEQ ID NO:140.  
 Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
 disease monitoring; cytokine; cell proliferation; cell differentiation;  
 immune modulation; haematopoiesis regulation; tissue growth;  
 angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
 thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 behaviour; cancer; proliferative disorder; neurological disorder;  
 cardiovascular disease; immune system disorder; organ transplantation;  
 tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 hypothyroidism; cholesterol ester storage disease; infection; vulvar;  
 vasotropic; antipsoriatic; antidiabetic; cytostatic; neutropenic;  
 neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
 cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
 dermatological; analgesic; virucide; antibacterial; fungicide.  
 Homo sapiens.  
 WO200190366-A2.  
 29-NOV-2001.  
 24-MAY-2001; 2001WO-US017076.  
 24-MAY-2000; 2000US-0206690P.  
 (CURA-) CURAGEN CORP.  
 Leach MD, Shinkets RA;  
 WPI; 2002-106200/14.  
 N-PSDB; ABN75123.  
 Novel human polypeptides and polynucleotides useful for diagnosing,  
 preventing and treating cardiovascular disease, neurodegenerative,  
 hyperproliferative disorders and disorders related to organ  
 transplantation.  
 Claim 10; Page 305; 2508pp; English.  
 Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 ABN79587 represent cDNAs encoding them. The invention also encompasses  
 polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 polynucleotides, the recombinant production of ORFX proteins, antibodies  
 specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 polypeptides, methods of screening for modulators of ORFX expression or  
 activity, and methods of screening individuals for a predisposition to an  
 ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 range of biological activities, such as cytokine, cell proliferation,  
 cell differentiation, immune modulation, haematopoiesis regulation,  
 tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 chemokinetic activity, haemostatic activity, thrombolytic activity,  
 receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 and antiinfective activity, and may also be involved in the determination  
 of bodily characteristics, fertility and behaviour. ORFX proteins,  
 nucleic acids and antibodies may be used in the treatment of cancers,  
 other proliferative disorders such as psoriasis and benign tumours,  
 neurological disorders such as epilepsy and Alzheimer's disease,  
 cardiovascular diseases, immune system disorders, disorders related to  
 organ transplantation, disorders of tissue growth and regeneration,  
 diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 storage disease, and infectious diseases caused by viral, bacterial,  
 fungal and other pathogens. ORFX nucleic acids may also be used as a  
 source of primers and probes, in the detection of ORFX genomic sequences  
 or transcripts, in the identification and cloning of homologous  
 sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 nucleic acids may additionally be used to produce transgenic animals  
 which may be useful for studying the function and/or activity of ORFX  
 protein, and in drug screening. The ORFX proteins may also be used as  
 immunogens to generate specific antibodies, which are useful in the  
 diagnosis, treatment and monitoring of ORFX-associated diseases  
 QY 1 PYKLELVGVQRLKYAQ 18  
 Query Match 45.6%; Score 46.5; DB 5; Length 65;  
 Best Local Similarity 55.6%; Pred. No. 6.4;  
 Matches 10; Conservative 3; Mismatches 4; Indels 1; Gaps 1;



Wed May 12 03:55:30 2021

Db 34 FYKGLR-EISRNLRLRYAQ 50

RESULT 28  
ABP65745  
ID ABP65745 standard; protein; 375 AA.  
XX AC  
XX ABP65745;  
XX DT 19-NOV-2002 (first entry)  
XX DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:499.  
XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
XX antidiarrhetic; antibacterial; inhibitor of Salmonella; detection;  
XX identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;  
XX rotavirus; food composition; pharmaceutical composition.  
XX Bifidobacterium longum.  
XX OS  
XX EP1227152-A1.  
XX PN  
XX 31-JUL-2002.  
XX PD  
XX 30-JAN-2001; 2001EP-00102050.  
XX PF  
XX 30-JAN-2001; 2001EP-00102050.  
XX PR  
XX (NEST ) SOC PROD NESTLE SA.  
XX PA  
XX WPI; 2002-668397/72.  
XX DR  
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as  
PT a probe or primer for detecting and/or identifying Bifidobacterium longum  
PT in a biological sample.  
XX PT  
XX Claim 3; SEQ ID NO 489; 80pp; English.  
XX PS  
XX The present invention describes a polynucleotide (I) comprising a  
CC sequence of a Bifidobacterium genome selected from the nucleotide  
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at  
CC least 90% identity or which hybridises with the sequences given in  
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a  
CC fusion protein, comprising a sequence selected from 1097 sequences given  
CC in ABP65258 to ABP65354 ligated in frame to a polynucleotide encoding a  
CC heterologous polypeptide. (I) has antidiarrhetic and antibacterial  
CC activities, and can be used as an inhibitor of Salmonella. (II) (which is  
CC a probe) is useful for the detection and/or identification of  
CC Bifidobacterium longum in a biological sample. A carrier containing the  
CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618) can be  
CC used for preventing and/or treating diarrhoea brought about by pathogenic  
CC bacteria and/or rotavirus. The carrier is a food composition selected  
CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented  
CC products, ice-creams, fermented cereal based products, milk based  
CC powders, infant formula, pet food or a pharmaceutical composition  
CC selected from tablets, liquid bacterial suspensions, dried oral  
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
CC (I) is useful in DNA arrays or chips to carry out analysis of the  
CC expression of the Bifidobacterium genome. ABQ81844 to ABQ81850 represent  
CC Bifidobacterium related nucleotide sequences given in the Sequence  
CC Listing from the present invention but not mentioned further within the  
CC specification. N.B. the sequence data for this patent is not represented  
CC in the printed specification but is based on sequence information  
CC supplied by the European Patent Office  
XX CC  
XX Sequence 375 AA;  
Query Match 45.1%; Score 46; DB 5; Length 375;  
Best Local Similarity 47.4%; Pred. No. 50;  
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
1 FYKGLR-EISRNLRLRYAQ 19

Db 68 FYAPLKMDPGASVLHYAQ 86

RESULT 29  
AAW42924  
ID AAW42924 standard; peptide; 20 AA.  
XX AC  
XX AAW42924;  
XX DT 28-APR-1998 (first entry)  
XX DE Immunogenic Hepatitis A virus peptide YK-1317.  
XX Immunogenic peptide; immunogenic epitope; P2A protein; immune response;  
XX antibody.  
XX KW  
XX Synthetic.  
XX OS  
XX Hepatitis A virus.  
XX PN WO9740147-A1.  
XX FN  
XX 30-OCT-1997.  
XX PD  
XX 18-APR-1997; 97WO-US006891.  
XX PF  
XX 19-APR-1996; 96US-0015644P.  
XX PR  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX PA  
XX Fields HA, Khudyakov YE;  
XX PI  
XX WPI; 1997-535831/49.  
XX DR  
XX Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
PT response to HAV in a mammal or to detect the presence of antibodies  
PT against HAV in a mammal.  
XX PT  
XX Claim 18; Page 112; 140pp; English.  
XX PS  
XX Peptides AAW42922-30 are immunogenic peptides corresponding to  
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
CC substantially similar to a portion of the amino acid sequence of the P2A  
CC protein of HAV corresponding to amino acids 792-980. The present peptide  
CC is derived from amino acids 810-829, and has a reactivity of 83.3% with  
CC acute sera. Compositions containing the peptides can be used to induce an  
CC immune response to HAV in a mammal. The peptides can also be used to  
CC detect the presence of antibodies against HAV in mammalian serum. The  
CC peptides can also be used to make an antibody against HAV by  
CC administering the peptide to a mammal  
XX CC  
XX Sequence 20 AA;  
Query Match 44.1%; Score 45; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 12 QRLKVAQEE 20  
DB 1 QRLKVAQEE 9  
RESULT 30  
AAB69441  
ID AAB69441 standard; peptide; 21 AA.  
XX AC  
XX AAB69441;  
XX AC  
XX 20-APR-2001 (first entry)  
XX DT  
XX Synthetic HAV P2A peptide, SEQ ID NO: 41.  
XX DE  
XX Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
KW

KW antigen; major structural capsid polypeptide; HAV antibody detection.

OS Hepatitis A virus.  
OS Synthetic.

PN WO200105824-A2.

XX 25-JAN-2001.

PD 14-JUL-2000; 2000WO-US019267.

XX 15-JUL-1999; 99US-0144412P.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX Fields HA, Khudyakov YE;

XX WPI; 2001-112681/12.

XX Synthetic peptides used as antigen sources for enzyme immunoassays  
PT detecting anti-hepatitis A virus and as vaccines.

XX Claim 13; Page 94; 130pp; English.

XX The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IGM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy end  
CC of the peptides enhances the IGM antibody reactivity

XX SQ Sequence 21 AA;

Query Match 44.1%; Score 45; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred.No. 3-4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 QRLKYAQEE 20

Db 1 QRLKYAQEE 9

Search completed: May 11, 2004, 13:37:25  
Job time : 51 secs

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:34:17 ; Search time 13.5 Seconds  
(without alignments)  
76.483 Million cell updates/sec

Title: US-09-171-432A-40

Perfect score: 102

Sequence: 1 PYKELRLEVKQRLKYAQEE 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*

2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*

3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep:\*

4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep:\*

5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep:\*

6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	102	100.0	1091	6	Patent No. 5516630-2
2	102	100.0	2227	3	Sequence 2, Appli
3	102	100.0	2227	3	Sequence 4, Appli
4	102	100.0	2227	3	Sequence 6, Appli
5	102	100.0	2227	3	Sequence 2, Appli
6	102	100.0	2227	3	Sequence 4, Appli
7	102	100.0	2227	3	Sequence 2, Appli
8	102	100.0	2227	4	Sequence 2, Appli
9	102	100.0	2227	4	Sequence 4, Appli
10	102	100.0	2227	4	Sequence 6, Appli
11	102	100.0	2227	4	Sequence 12, Appli
12	102	100.0	2227	4	Sequence 2, Appli
13	102	100.0	2227	4	Sequence 4, Appli
14	102	100.0	2227	4	Sequence 6, Appli
15	94	92.2	839	1	Sequence 2, Appli
16	47	46.1	249	4	Sequence 108, App
17	45	44.1	214	4	Sequence 4748, Ap
18	43	42.2	50	4	Sequence 4556, Ap
19	42	41.2	452	4	Sequence 28627, A
20	42	41.2	789	4	Sequence 9210, Ap
21	42	41.2	2190	4	Sequence 25754, A
22	41	40.2	60	4	Sequence 4371, Ap
23	41	40.2	68	1	Sequence 2, Appli
24	41	40.2	68	1	Sequence 4, Appli
25	41	40.2	68	2	Sequence 2, Appli
26	41	40.2	68	2	Sequence 4, Appli
27	41	40.2	69	5	Sequence 5, Appli

28	41	40.2	69	5	PCT-US95-06406A-7	Sequence 7, Appli
29	41	40.2	147	1	US-08-364-003B-2	Sequence 2, Appli
30	41	40.2	147	2	US-08-959-865-3	Sequence 3, Appli
31	41	40.2	147	3	US-08-842-234-2	Sequence 2, Appli
32	41	40.2	147	4	US-09-671-317-488	Sequence 488, App
33	41	40.2	189	4	US-09-134-000C-4845	Sequence 4845, Ap
34	41	40.2	534	4	US-09-252-991A-32086	Sequence 32086, A
35	41	40.2	794	4	US-09-252-991A-29088	Sequence 29088, A
36	41	40.2	796	4	US-09-252-991A-17763	Sequence 17763, A
37	40	39.2	182	4	US-09-439-261-21	Sequence 21, Appl
38	40	39.2	182	4	US-09-227-613-20	Sequence 20, Appl
39	40	39.2	189	4	US-09-080-643-2	Sequence 2, Appli
40	40	39.2	189	4	US-09-080-643-4	Sequence 4, Appli
41	40	39.2	219	4	US-09-439-261-20	Sequence 20, Appl
42	40	39.2	219	4	US-09-227-613-19	Sequence 19, Appl
43	40	39.2	287	4	US-09-439-261-13	Sequence 13, Appl
44	40	39.2	287	4	US-09-227-613-14	Sequence 14, Appl
45	40	39.2	288	4	US-09-439-261-14	Sequence 14, Appl
46	40	39.2	288	4	US-09-439-261-16	Sequence 16, Appl
47	40	39.2	288	4	US-09-439-261-18	Sequence 18, Appl
48	40	39.2	288	4	US-09-227-613-15	Sequence 15, Appl
49	40	39.2	347	4	US-09-439-261-42	Sequence 42, Appl
50	40	39.2	347	4	US-09-227-613-40	Sequence 40, Appl
51	40	39.2	360	4	US-09-439-261-41	Sequence 41, Appl
52	40	39.2	360	4	US-09-227-613-39	Sequence 39, Appl
53	40	39.2	444	4	US-09-439-261-11	Sequence 11, Appl
54	40	39.2	444	4	US-09-439-261-43	Sequence 43, Appl
55	40	39.2	444	4	US-09-227-613-12	Sequence 12, Appl
56	40	39.2	444	4	US-09-227-613-42	Sequence 42, Appl
57	40	39.2	444	4	US-09-048-888-3	Sequence 3, Appli
58	40	39.2	445	4	US-09-439-261-39	Sequence 39, Appl
59	40	39.2	445	4	US-09-439-261-45	Sequence 45, Appl
60	40	39.2	487	4	US-09-252-991A-28079	Sequence 28079, A
61	40	39.2	498	4	US-09-198-452A-143	Sequence 143, App
62	40	39.2	586	4	US-09-252-991A-27398	Sequence 27398, A
63	40	39.2	764	4	US-09-370-838-67	Sequence 67, Appl
64	40	39.2	1463	1	US-08-157-005-3	Sequence 3, Appli
65	40	39.2	1463	3	US-08-747-863-3	Sequence 3, Appli
66	40	39.2	1463	4	US-09-565-864-3	Sequence 3, Appli
67	39.5	38.7	384	4	US-09-310-363C-2	Sequence 2, Appli
68	39	38.2	183	4	US-08-477-831C-40	Sequence 40, Appl
69	39	38.2	244	4	US-09-198-432A-723	Sequence 723, App
70	39	38.2	347	4	US-09-134-000C-3690	Sequence 3690, Ap
71	39	38.2	409	4	US-09-310-363C-4	Sequence 4, Appli
72	39	38.2	1408	1	US-08-612-521-2	Sequence 2, Appli
73	39	38.2	1898	1	US-08-056-200-94	Sequence 94, Appl
74	39	38.2	1898	4	US-08-800-644-94	Sequence 94, Appl
75	38.5	37.7	170	4	US-08-529-055-32	Sequence 32, Appl
76	38.5	37.7	283	4	US-09-328-352-6473	Sequence 6473, Ap
77	38.5	37.7	641	3	US-08-961-083-160	Sequence 160, App
78	38.5	37.7	641	4	US-09-536-784-160	Sequence 160, App
79	38.5	37.7	1007	4	US-09-957-005-9	Sequence 9, Appli
80	38.5	37.7	8991	4	US-08-714-741-32	Sequence 32, Appl
81	38	37.3	128	4	US-09-134-001C-3984	Sequence 3984, Ap
82	38	37.3	129	4	US-08-981-527A-11	Sequence 11, Appl
83	38	37.3	129	4	US-09-489-039A-11444	Sequence 11444, A
84	38	37.3	211	4	US-09-134-001C-4162	Sequence 4162, Ap
85	38	37.3	233	3	US-08-836-236-8	Sequence 8, Appli
86	38	37.3	256	3	US-08-719-758-2	Sequence 2, Appli
87	38	37.3	256	3	US-09-113-827-2	Sequence 2, Appli
88	38	37.3	261	4	US-09-252-991A-27270	Sequence 27270, A
89	38	37.3	391	4	US-09-543-681A-4956	Sequence 4956, Ap
90	38	37.3	584	4	US-09-252-991A-19593	Sequence 19593, A
91	38	37.3	584	4	US-09-252-991A-18292	Sequence 18292, A
92	38	37.3	1050	4	US-09-555-554-2	Sequence 2, Appli
93	38	37.3	3072	3	US-09-413-814-93	Sequence 93, Appl
94	38	37.3	3079	3	US-09-413-814-80	Sequence 80, Appl
95	37.5	36.8	18	3	US-08-940-095-229	Sequence 229, App
96	37.5	36.8	18	3	US-08-940-093-229	Sequence 229, App
97	37.5	36.8	18	3	US-08-940-096-229	Sequence 229, App
98	37.5	36.8	18	3	US-09-465-719-229	Sequence 229, App
99	37.5	36.8	18	4	US-09-453-605-229	Sequence 229, App
100	37.5	36.8	18	4	US-09-453-638-229	Sequence 229, App

```

; Patent No. 6113912
;
; GENERAL INFORMATION:
;
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
;
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
;
; FILE REFERENCE: 20264262US2
; CURRENT APPLICATION NUMBER: US/08/475,886A
; CURRENT FILING DATE: 1995-06-07
;
; EARLIER APPLICATION NUMBER: 07/947,338
; EARLIER FILING DATE: 1992-09-18
; EARLIER APPLICATION NUMBER: 08/397,232
; EARLIER FILING DATE: 1995-03-10
;
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 4
;
; LENGTH: 2227
;
; TYPE: PRT
;
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
;
; US-08-475-886-4

```

```
Query Match      100.0%; Score 102; DB 3; Length 2227;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	PKYELRLEVGKQRLKYAQEE	20
Db	799	PKYELRLEVGKQRLKYAQEE	818

RESULT 4  
US-08-475-886-6  
: Sequence 6, Application US/08475886A

; GENERAL INFORMATION:  
 ; APPLICANT: FUNKHOUSER, ANN W  
 ; APPLICANT: EMERSON, SUZANNE U

APPLICANT: D'HONDT, ERIC  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
FILE REFERENCE: 20264262US2

, CURRENT FILING DATE: 1995-06-07  
 ,  
 , EARLIER APPLICATION NUMBER: 07/947,338  
 ,  
 , EARLIER FILING DATE: 1992-09-18  
 ,

```

; EARLIER FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent Ver. 2.1

```

```

: SER ID NO 8
:
: LENGTH: 2227
:
: TYPE: PRT
:
: ORGANISM: Attenuated (4380) HAV, strain HM-175
:

```

Query Match	100.0%;	Score 102;	DB 3;	Length 2227;
Best Local Similarity	100.0%;	Pred. No. 4e-08;		

1 PYKELELVGKORLKYAQEE 20

DD	TIME	IN	OUT	AVG	MAX	MIN	STDEV	COV
RESULT 5								

Sequence 2, Application US/08397232A  
Patent No. 6180110  
GENERAL INFORMATION:

APPLICANT: EMERSON, SUZANNE U  
APPLICANT: PURCELL, ROBERT H

APPLICANT: D'HONDT, ERIC  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
FILE REFERENCE: 20264262US1  
CURRENT APPLICATION NUMBER: US/08/397,232A  
CURRENT FILING DATE: 1995-04-17  
EARLIER APPLICATION NUMBER: 07/947,338  
EARLIER FILING DATE: 1992-09-18  
EARLIER APPLICATION NUMBER: PCT/US93/08610  
EARLIER FILING DATE: 1993-09-17  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-08-397-232-2

Query Match 100.0%; Score 102; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 799 PYKELRLEVGKQRLKYAQEE 818

RESULT 6  
US-08-397-232-4  
Sequence 4, Application US/08397232A  
Patent No. 6180110  
GENERAL INFORMATION:  
APPLICANT: FUNKHOUSER, ANN W  
APPLICANT: EMERSON, SUZANNE U  
APPLICANT: PURCELL, ROBERT H  
APPLICANT: D'HONDT, ERIC  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
FILE REFERENCE: 20264262US1  
CURRENT APPLICATION NUMBER: US/08/397,232A  
CURRENT FILING DATE: 1995-04-17  
EARLIER APPLICATION NUMBER: 07/947,338  
EARLIER FILING DATE: 1992-09-18  
EARLIER APPLICATION NUMBER: PCT/US93/08610  
EARLIER FILING DATE: 1993-09-17  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-08-397-232-4

Query Match 100.0%; Score 102; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 799 PYKELRLEVGKQRLKYAQEE 818

RESULT 7  
US-09-171-387-2  
Sequence 2, Application US/09171387  
Patent No. 6280734  
GENERAL INFORMATION:  
APPLICANT: RAYCHAUDHURI, GOPA;  
EMERSON, SUZANNE, U.;  
PURCELL, ROBERT, H.  
TITLE OF INVENTION: SIMIAN-HUMAN HAV  
HAVING A CHIMERIC 2C PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.

STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MICROSOFT WORD 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/171,387  
FILING DATE: 24-Mar-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/06506  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US60/015,642  
FILING DATE: 19-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: William S. Feiler  
REGISTRATION NUMBER: 26,728  
REFERENCE/DOCKET NUMBER: 2026-4229US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
INFORMATION FOR SEQ ID NO: 2  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2227 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-171-387-2

Query Match 100.0%; Score 102; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 799 PYKELRLEVGKQRLKYAQEE 818

RESULT 8  
US-09-653-499-2  
Sequence 2, Application US/09653499  
Patent No. 6423318  
GENERAL INFORMATION:  
APPLICANT: FUNKHOUSER, ANN W  
APPLICANT: EMERSON, SUZANNE U  
APPLICANT: PURCELL, ROBERT H  
APPLICANT: D'HONDT, ERIC  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
FILE REFERENCE: 20264262US2  
CURRENT APPLICATION NUMBER: US/09/653,499  
CURRENT FILING DATE: 2000-08-31  
PRIOR APPLICATION NUMBER: 08/475,886  
PRIOR FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: 07/947,338  
PRIOR FILING DATE: 1992-09-18  
PRIOR APPLICATION NUMBER: 08/397,232  
PRIOR FILING DATE: 1995-03-10  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-09-653-499-2

Query Match 100.0%; Score 102; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKEIRLEVGKORLYAQEE 20  
 |||||  
 799 PYKEIRLEVGKORLYAQEE 818

RESULT 11

```

US-10-104-966-12
; Sequence 12, Application US/10104966
; Patent NO. 6680059
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis A virus sequence
US-10-104-966-12

```

```

; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-09-653-499-4

Query Match      100.0%; Score 102; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Query Match	100.0%;	Score 102;	DB 4;	Length 2227;
Best Local Similarity	100.0%;	Pred. No. 4e-08;		
Mismatches	0;	Mismatches	0;	Indels 0;
Conservative	20;	Conservative	20;	Indels 0;

```

Oy      1 PYKELRLEVKGQRLKYAOEE 20
          |||||
nh      799 PYKELRLEVKGQRLKYAOEE 818
          |||||
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 PYKELRLEVKGQRLKYAQEE 20  
799 PYKELRLEVKGQRLKYAQEE 818

## RESULT 12

```

US-10-135-988-2
; Sequence 2, Application US/10135988
; Patent No. 6680060
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US3
; CURRENT APPLICATION NUMBER: US/10/135,988
; CURRENT FILING DATE: 2002-04-29
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2227
; TYPE: PR1
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175
US-10-135-988-2

```

```

; LENGTH: 2227
; TYPE: PR
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-09-653-499-6

Query Match      Score 102; DB 4; Length 2227;
Best local similarity 100.0%; Pred. No. 4e-08;
Identical 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
Query Match      100.0%; Score 102; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0
```

Query Match	100.0%	Score 102;	DB 4;	Length 2227;
Best Local Similarity	100.0%;	Pred. No. 48-08;		
Misbase %	20.0	Mismatches 0;	Indels 0;	Gaps 0;
Conservative				

QY 1 PYKELRLEVGVGKQRLKYAQEE 20  
799 PVKEIRLEVGVGKQRLKYAQEE 818  
nb

```
RESULT 13
US-10-135-988-4
; Sequence 4, Application US/10135988
; Patent No. 6680060
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US3
; CURRENT APPLICATION NUMBER: US/10/135,988
; CURRENT FILING DATE: 2002-04-29
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-10-135-988-4
Query Match 100.0%; Score 102; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELELVGKQRLKYAEE 20
DB 799 PYKELELVGKQRLKYAEE 818

RESULT 14
US-10-135-988-6
; Sequence 6, Application US/10135988
; Patent No. 6680060
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US3
; CURRENT APPLICATION NUMBER: US/10/135,988
; CURRENT FILING DATE: 2002-04-29
; PRIOR APPLICATION NUMBER: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-10-135-988-6
Query Match 100.0%; Score 102; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELELVGKQRLKYAEE 20
DB 799 PYKELELVGKQRLKYAEE 818

RESULT 15
US-08-087-016-2
; Sequence 2, Application US/08087016
; Patent No. 5430135
; GENERAL INFORMATION:
; APPLICANT: NAINAN, OMAYA V.
; APPLICANT: MARGOLIS, HAROLD S.
; APPLICANT: ROBERTSON, BETTY H.
; APPLICANT: BRINTON, MARGO H.
; APPLICANT: EBERT, JAMES W.
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1615 L Street N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,016
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,828
; FILING DATE: 03-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26,581
; REFERENCE/DOCKET NUMBER: WTS/5683/83834/SRL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 839 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-087-016-2
Query Match 92.2%; Score 94; DB 1; Length 839;
Best Local Similarity 90.0%; Pred. No. 2.9e-07;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PYKELELVGKQRLKYAEE 20
DB 798 PYKELELVGKQRLKYAEE 817

RESULT 16
US-09-328-352-4748
; Sequence 4748, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4748
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4748
Query Match 46.1%; Score 47; DB 4; Length 249;
Best Local Similarity 56.2%; Pred. No. 4;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```





; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.

US-09-252-991A-25754

Query Match 41.2%; Score 42; DB 4; Length 2190;  
Best Local Similarity 47.4%; Pred. No. 2.6e-02;  
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 PYKELRLEVGVKQRLKYAQ 19  
DB 40 PIPQTRQEMGNPLSLVAQE 58

RESULT 22

US-09-621-976-4371  
; Sequence 4371, Application US/09621976  
; Patent No. 6639083  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.

; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 4371  
; LENGTH: 60

; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: 8  
; OTHER INFORMATION: Xaa = Arg,Thr  
; NAME/KEY: UNSURE  
; LOCATION: 9  
; OTHER INFORMATION: Xaa = His,Leu,Pro,Arg

US-09-621-976-4371

Query Match 40.2%; Score 41; DB 4; Length 60;  
Best Local Similarity 43.8%; Pred. No. 8.7;  
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 KELRLEVGVKQRLKYAQ 18  
DB 15 EQKLEAGVERIKVQSQ 30

RESULT 23

US-08-606-789-2  
; Sequence 2, Application US/08606789  
; Patent No. 5783418

; GENERAL INFORMATION:

; APPLICANT: Au-Young, Janice  
; APPLICANT: Stuart, Susan G.  
; APPLICANT: Murry, Lynn E.  
; APPLICANT: Guegler, Kark J.  
; APPLICANT: Seilhamer, Jeffrey J.

; TITLE OF INVENTION: HUMAN HOMOLOG OF THE RAT G PROTEIN  
; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 1.5

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/606,789  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA: 08/440,743  
; APPLICATION NUMBER: 08/440,743  
; FILING DATE: May 5, 1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/320,011  
; FILING DATE: October 5, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Luther, Barbara J.  
; REGISTRATION NUMBER: 33,954  
; REFERENCE/DOCKET NUMBER: PF-0055 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-852-0195  
; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 68 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: PITUITARY  
; CLONE: 112530

US-08-606-789-2

Query Match 40.2%; Score 41; DB 1; Length 68;  
Best Local Similarity 50.0%; Pred. No. 9.9;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 KELRLEVGVKQRLKYAQ 18  
DB 15 QQLRLEAGLNRVKVSQ 30

RESULT 24

US-08-606-789-4

; Sequence 4, Application US/08606789  
; Patent No. 5783418

; GENERAL INFORMATION:

; APPLICANT: Au-Young, Janice  
; APPLICANT: Stuart, Susan G.  
; APPLICANT: Murry, Lynn E.  
; APPLICANT: Guegler, Kark J.  
; APPLICANT: Seilhamer, Jeffrey J.

; TITLE OF INVENTION: HUMAN HOMOLOG OF THE RAT G PROTEIN

; TITLE OF INVENTION: GAMMA-5 SUBUNIT

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/606,789

; FILING DATE: Filed Herewith

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/440,743

; FILING DATE: May 5, 1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/320,011

/ FILING DATE: October 5, 1994  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Luther, Barbara J.  
/ REGISTRATION NUMBER: 33,954  
/ REFERENCE/DOCKET NUMBER: PF-0055 US  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 415-855-0555  
/ TELEFAX: 415-852-0195  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 4:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 68 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: peptide  
/ IMMEDIATE SOURCE:  
/ LIBRARY: GenBank  
/ CLONE: 204241  
/ US-08-606-789-4

Query Match 40.2%; Score 41; DB 1; Length 68;  
Best Local Similarity 50.0%; Pred. No. 9.9;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 KELRLGVKQRLKYAQ 18  
Db 15 QQLRLKGLNLRVKVQSQ 30

RESULT 25  
US-09-111-348-2  
/ Sequence 2, Application US/09111348  
/ Patent No. 5912130  
/ GENERAL INFORMATION:  
/ APPLICANT: Au-Young, Janice  
/ APPLICANT: Stuart, Susan G.  
/ APPLICANT: Murry, Lynn E.  
/ APPLICANT: Guegler, Kark J.  
/ APPLICANT: Seilhamer, Jeffrey J.  
/ TITLE OF INVENTION: HUMAN HOMOLOG OF THE RAT G PROTEIN  
/ NUMBER OF SEQUENCES: 31  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Incyte Pharmaceuticals, Inc.  
/ STREET: 3174 Porter Drive  
/ CITY: Palo Alto  
/ STATE: CA  
/ COUNTRY: U.S.  
/ ZIP: 94303  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Diskette  
/ COMPUTER: IBM Compatible  
/ OPERATING SYSTEM: DOS  
/ SOFTWARE: FASTSEQ Version 1.5  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/09/111,348  
/ FILING DATE:  
/ CLASSIFICATION:  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/606,789  
/ FILING DATE:  
/ APPLICATION NUMBER: 08/440,743  
/ FILING DATE: May 5, 1995  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/320,011  
/ FILING DATE: October 5, 1994  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Luther, Barbara J.  
/ REGISTRATION NUMBER: 33,954  
/ REFERENCE/DOCKET NUMBER: PF-0055 US  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 415-855-0555

/ TELEFAX: 415-852-0195  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 2:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 68 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: peptide  
/ IMMEDIATE SOURCE:  
/ LIBRARY: PITUITARY  
/ CLONE: 112530  
/ US-09-111-348-2

Query Match 40.2%; Score 41; DB 2; Length 68;  
Best Local Similarity 50.0%; Pred. No. 9.9;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 KELRLGVKQRLKYAQ 18  
Db 15 QQLRLKGLNLRVKVQSQ 30

RESULT 26  
US-09-111-348-4  
/ Sequence 4, Application US/09111348  
/ Patent No. 5912130  
/ GENERAL INFORMATION:  
/ APPLICANT: Au-Young, Janice  
/ APPLICANT: Stuart, Susan G.  
/ APPLICANT: Murry, Lynn E.  
/ APPLICANT: Guegler, Kark J.  
/ APPLICANT: Seilhamer, Jeffrey J.  
/ TITLE OF INVENTION: HUMAN HOMOLOG OF THE RAT G PROTEIN  
/ NUMBER OF SEQUENCES: 31  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Incyte Pharmaceuticals, Inc.  
/ STREET: 3174 Porter Drive  
/ CITY: Palo Alto  
/ STATE: CA  
/ COUNTRY: U.S.  
/ ZIP: 94303  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Diskette  
/ COMPUTER: IBM Compatible  
/ OPERATING SYSTEM: DOS  
/ SOFTWARE: FASTSEQ Version 1.5  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/09/111,348  
/ FILING DATE:  
/ CLASSIFICATION:  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/606,789  
/ FILING DATE:  
/ APPLICATION NUMBER: 08/440,743  
/ FILING DATE: May 5, 1995  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/320,011  
/ FILING DATE: October 5, 1994  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Luther, Barbara J.  
/ REGISTRATION NUMBER: 33,954  
/ REFERENCE/DOCKET NUMBER: PF-0055 US  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 415-855-0555  
/ TELEFAX: 415-852-0195  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 4:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 68 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 204241
US-09-111-348-4

Query Match 40.2%; Score 41; DB 2; Length 68;
Best Local Similarity 50.0%; Pred. No. 9.9;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 KELRLVGVKQRLKYAQ 18
Db 15 QQLRLAAGLRVVKVQSQ 30

RESULT 27
PCT-US95-06406A-5
; Sequence 5, Application PC/TUS9506406A
; GENERAL INFORMATION:
; APPLICANT: Janet D. Robishaw, Charles Kunsch
; TITLE OF INVENTION: cDNA Clones Encoding Human G Protein
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06406A
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 23
; SEQUENCE CHARACTERISTICS:
; LENGTH: 69
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; PCT-US95-06406A-5

Query Match 40.2%; Score 41; DB 2; Length 68;
Best Local Similarity 50.0%; Pred. No. 9.9;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 KELRLVGVKQRLKYAQ 18
Db 15 QQLRLAAGLRVVKVQSQ 30

RESULT 28
PCT-US95-06406A-7
; Sequence 7, Application PC/TUS9506406A
; GENERAL INFORMATION:
; APPLICANT: Janet D. Robishaw, Charles Kunsch
; TITLE OF INVENTION: cDNA Clones Encoding Human G Protein
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06406A
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 69
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; PCT-US95-06406A-5

Query Match 40.2%; Score 41; DB 5; Length 69;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 KELRLVGVKQRLKYAQ 18
Db 15 QQLRLAAGLRVVKVQSQ 30

RESULT 29
US-08-264-003B-2
; Sequence 2, Application US/08264003B
; Patent No. 5698076
; GENERAL INFORMATION:
; APPLICANT: Gentz, Reiner
; APPLICANT: Fleischmann, Robert D.
; TITLE OF INVENTION: 5-LIPOXYGENASE-ACTIVATING PROTEIN II
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
; ADDRESSEE: STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,003B
; FILING DATE: 06-JUN-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-117
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 147 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-264-003B-2

Query Match 40.2%; Score 41; DB 1; Length 147;  
Best Local Similarity 80.0%; Pred. No. 22;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 7 LEVGKRLKY 16  
Db 24 LQVGKRLKY 33

RESULT 30  
US-08-959-865-3  
Sequence 3, Application US/08959865  
Patent No. 5919627  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: MICROSOMAL GLUTATHIONE-S TRANSFERASE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/959,865  
FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0411 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 147 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: COLN0707  
CLONE: 903729  
US-08-959-865-3

Query Match 40.2%; Score 41; DB 2; Length 147;  
Best Local Similarity 80.0%; Pred. No. 22;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 7 LEVGKRLKY 16

Db 24 LQVGKRLKY 33

Search completed: May 11, 2004, 13:42:20  
Job time : 14.5 secs

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:30:01 ; Search time 10.5 Seconds  
(without alignments)  
183.222 Million cell updates/sec

Title: US-09-171-432A-40

Perfect score: 102

Sequence: 1 PYKELRLEVGKQLKYAQEE 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR\_78:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description
1	102	100.0	55	2	P00433	genome polyprotein
2	102	100.0	56	2	P00434	genome polyprotein
3	102	100.0	56	2	P00428	genome polyprotein
4	102	100.0	56	2	P00432	genome polyprotein
5	102	100.0	56	2	P00430	genome polyprotein
6	102	100.0	341	2	S04137	genome polyprotein
7	102	100.0	852	1	GNNYHA	genome polyprotein
8	102	100.0	1358	2	A03905	genome polyprotein
9	102	100.0	2227	1	GNNYHM	genome polyprotein
10	102	100.0	2227	1	GNNYHR	genome polyprotein
11	102	100.0	2227	1	GNNYMK	genome polyprotein
12	102	100.0	2227	1	GNNYHE	genome polyprotein
13	100	98.0	56	2	P00427	genome polyprotein
14	99	97.1	56	2	P00429	genome polyprotein
15	95	93.1	56	2	P00436	genome polyprotein
16	95	93.1	2230	1	GNNYSA	genome polyprotein
17	94	92.2	55	2	P00435	genome polyprotein
18	94	92.2	56	2	P00437	genome polyprotein
19	94	92.2	839	1	GNNYS2	genome polyprotein
20	93	91.2	56	2	P00431	genome polyprotein
21	62	60.8	319	2	JH0135	genome polyprotein
22	46	45.1	219	2	T45997	hypothetical prote
23	46	45.1	285	2	T29804	hypothetical prote
24	45	44.1	209	2	D7121	ATP synthase chain
25	45	44.1	209	2	B86502	ATP synthase subun
26	44	43.1	179	2	AI0819	probable exported
27	44	43.1	290	2	C71701	branched-chain ami
28	44	43.1	776	2	T05938	hypothetical prote
29	43	42.2	207	2	H85609	hypothetical prote

30	43	42.2	207	2	A99801	hypothetical prote
31	43	42.2	208	1	F70485	uracil phosphoribo
32	43	42.2	465	2	T25441	hypothetical prote
33	43	42.2	509	2	G81929	probable iron-upta
34	43	42.2	534	2	JC5096	transposase - fung
35	43	42.2	3587	2	T31075	tyrocidine synthet
36	42.5	41.7	179	2	H91049	probable membrane
37	42.5	41.7	179	2	E85894	hypothetical prote
38	42.5	41.7	179	2	A65027	ATP-dependent heli
39	42.5	41.7	722	2	A02222	hypothetical prote
40	42	41.2	290	2	B97774	probable cysteine
41	42	41.2	365	2	F86413	probable porin PA0
42	42	41.2	452	2	B83623	flagellar biosynth
43	42	41.2	695	2	AH2647	flagellar biosynth
44	42	41.2	723	2	G97429	ribose/galactose A
45	42	41.2	736	2	B82944	DNA topoisomerase
46	42	41.2	752	2	A45582	DNA topoisomerase
47	42	41.2	752	2	A65089	DNA topoisomerase
48	42	41.2	752	2	G91116	DNA topoisomerase
49	42	41.2	752	2	G85961	topoisomerase IV c
50	42	41.2	752	2	AF0888	hypothetical prote
51	42	41.2	826	2	H84683	valine-tRNA ligase
52	42	41.2	939	2	H71532	protein T6D22.8 (i
53	42	41.2	990	2	A86215	hypothetical prote
54	42	41.2	4131	2	T21085	GTP-binding regula
55	41	40.2	68	2	B42243	GTP-binding regula
56	41	40.2	68	2	I39158	conserved hypotet
57	41	40.2	151	2	A72409	hypothetical prote
58	41	40.2	384	2	E84188	aminopeptidase II
59	41	40.2	412	2	E70108	probable atpH prot
60	41	40.2	446	2	G70774	hypothetical prote
61	41	40.2	515	2	H75579	ATP-dependent RNA
62	41	40.2	585	2	G96995	hypothetical prote
63	41	40.2	635	2	D84920	hypothetical prote
64	41	40.2	656	2	E75468	hypothetical prote
65	41	40.2	735	2	E83540	conserved hypotet
66	41	40.2	777	2	B95991	heat shock atp-dep
67	41	40.2	842	2	G90576	ClpB protein NMA16
68	41	40.2	859	2	F81863	ClpB protein NMB14
69	41	40.2	859	2	F81078	RNA-directed RNA p
70	41	40.2	1116	2	T30828	hypothetical prote
71	41	40.2	1206	2	D90085	molybdenum-pterin-
72	40	39.2	69	2	I64119	hypothetical prote
73	40	39.2	178	2	F87408	peptidyl-tRNA hyd
74	40	39.2	189	2	E95000	aminoacyl-tRNA hyd
75	40	39.2	189	2	E97872	uridine kinase (EC
76	40	39.2	213	2	AH0185	DNA-binding respon
77	40	39.2	218	2	C82378	hypothetical prote
78	40	39.2	241	2	A71020	hypothetical prote
79	40	39.2	353	2	F69258	iron-sulfur bindin
80	40	39.2	366	2	B69313	translation elonga
81	40	39.2	456	2	S11665	BFR1 protein - Yea
82	40	39.2	470	2	S47887	hypothetical prote
83	40	39.2	485	2	E86506	hypothetical prote
84	40	39.2	485	2	E72115	hypothetical prote
85	40	39.2	485	2	A81555	hypothetical prote
86	40	39.2	501	2	C71948	RNA-directed RNA p
87	40	39.2	533	2	A45392	flagellar biosynth
88	40	39.2	583	2	A83530	methionine-tRNA ii
89	40	39.2	723	2	D71091	autocatalytic NOR-90
90	40	39.2	727	2	S18193	ribosomal transcri
91	40	39.2	727	2	JC5113	transcription fact
92	40	39.2	727	2	B40439	transcription fact
93	40	39.2	764	2	S03318	ribosomal transcri
94	40	39.2	764	2	JC5112	transcription fact
95	40	39.2	764	2	A40439	transcription fact
96	40	39.2	765	2	S22314	endopeptidase Ia (
97	40	39.2	795	1	S73830	outer arm dynein i
98	40	39.2	837	2	T02761	valyl-tRNA synthet
99	40	39.2	939	2	H81686	ori 1b protein - L
100	40	39.2	1463	2	A36861	

## ALIGNMENTS

RESULT 1  
PQ0433  
genome polyprotein - human hepatitis A virus (strain PA21) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0433  
A:Molecule type: mRNA  
A:Residues: 1-55 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 100.0%; Score 102; DB 2; Length 55;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 35 PYKELRLEVGKQRLKYAQEE 54

RESULT 2  
PQ0434  
genome polyprotein - human hepatitis A virus (strain KPH) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0434  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 100.0%; Score 102; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 35 PYKELRLEVGKQRLKYAQEE 54

RESULT 3  
PQ0428  
genome polyprotein - human hepatitis A virus (strain TKM002) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0428  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 100.0%; Score 102; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 36 PYKELRLEVGKQRLKYAQEE 55

RESULT 4  
PQ0432  
genome polyprotein - human hepatitis A virus (strain CF-53) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0432  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 100.0%; Score 102; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 36 PYKELRLEVGKQRLKYAQEE 55

RESULT 5  
PQ0430  
genome polyprotein - human hepatitis A virus (strain S23-1) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0430  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 100.0%; Score 102; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 36 PYKELRLEVGKQRLKYAQEE 55

RESULT 6  
PQ04137  
genome polyprotein - human hepatitis A virus (strain LDCG-1) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
R:Andonov, A.P.; Lau, P.C.K.; Chaudhary, R.  
Nucleic Acids Res. 17, 3594, 1989  
A:Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A virus  
A:Reference number: S04137; MUID:89263805; PMID:2542903  
A:Accession: S04137  
A:Molecule type: mRNA  
A:Residues: 1-341 <AND>

Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 36 PYKELRLEVGKQRLKYAQEE 55

## RESULT 4

PQ0432  
genome polyprotein - human hepatitis A virus (strain CF-53) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0432  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 100.0%; Score 102; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 36 PYKELRLEVGKQRLKYAQEE 55

## RESULT 5

PQ0430  
genome polyprotein - human hepatitis A virus (strain S23-1) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0430  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 100.0%; Score 102; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 1.8e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 36 PYKELRLEVGKQRLKYAQEE 55

## RESULT 6

S04137  
genome polyprotein - human hepatitis A virus (strain LDCG-1) (fragment)  
C:Species: human hepatitis A virus  
C>Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
R:Andonov, A.P.; Lau, P.C.K.; Chaudhary, R.  
Nucleic Acids Res. 17, 3594, 1989  
A:Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A virus  
A:Reference number: S04137; MUID:89263805; PMID:2542903  
A:Accession: S04137  
A:Molecule type: mRNA  
A:Residues: 1-341 <AND>

A;Cross-references: EMBL:X14666; NID:G62301; PIDN:CAA32794.1; PID:94377576

C;Genetics:

A;Gene: VP1

C;Superfamily: hepatitis A virus genome polyprotein

C;Keywords: coat protein; polyprotein

F;2-340/Product: coat protein 1D (VP1) #status predicted <MAT>

Query Match 100.0%; Score 102; DB 2; Length 341;

Best Local Similarity 100.0%; Pred. No. 1.3e-08; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0;

QY 1 PYKELRLEVGKQRLKYAOEE 20

|||||

DB 303 PYKELRLEVGKQRLKYAOEE 322

#### RESULT 7

GNVYHA

genome polyprotein - human hepatitis A virus (strain CR326) (fragment)

N;Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A

C;Species: human hepatitis A virus

A;Note: host Homo sapiens (man)

C;Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999

C;Accession: A03904

R;Linemeyer, D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitra, S.M.

J. Virol. 54, 247-255, 1985

A;Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.

A;Reference number: A03904; MUID:85185648; PMID:2985793

A;Accession: A03904

A;Molecule type: genomic RNA

A;Residues: 1-852 <LIN>

A;Cross-references: EMBL:M10033; NID:G329592; PIDN:AAA45470.1; PID:G329593

C;Superfamily: hepatitis A virus genome polyprotein

C;Keywords: coat protein; core protein; polyprotein

F;1-245/Product: coat protein 1A #status predicted <C1A>

F;246-491/Product: coat protein 1B #status predicted <C1B>

F;492-836/Product: coat protein 1C #status predicted <C1C>

F;837-852/Product: core protein 2A (fragment) #status predicted <C2A>

Query Match 100.0%; Score 102; DB 1; Length 852;

Best Local Similarity 100.0%; Pred. No. 3.7e-08; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0;

QY 1 PYKELRLEVGKQRLKYAOEE 20

|||||

DB 799 PYKELRLEVGKQRLKYAOEE 818

#### RESULT 8

A03905

genome polyprotein (version 2) - human hepatitis A virus (fragments)

N;Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; protein

C;Species: human hepatitis A virus

C;Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 15-Nov-1996

C;Accession: A03905

R;Baroudy, B.M.; Ticehurst, J.R.; Miele, T.A.; Maizel Jr., J.V.; Purcell, R.H.; Feinston

Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985

A;Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA

A;Reference number: A03905; MUID:85166289; PMID:2984684

A;Accession: A03905

A;Molecule type: genomic RNA

A;Residues: 1-1358 <BAR>

C;Superfamily: hepatitis A virus genome polyprotein

C;Keywords: coat protein; core protein; polyprotein

F;1-245/Product: coat protein 1A #status predicted <C1A>

F;246-491/Product: coat protein 1B #status predicted <C1B>

F;492-836/Product: coat protein 1C #status predicted <C1C>

F;837-854/Product: core protein 2A (fragment) #status predicted <C2A>

F;855-1358/Product: protein 3D (RNA polymerase) (fragment) #status predicted <C3D>

Query Match 100.0%; Score 102; DB 2; Length 1358;

Best Local Similarity 100.0%; Pred. No. 6.1e-08; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0;

QY 1 PYKELRLEVGKQRLKYAOEE 20

|||||

DB 799 PYKELRLEVGKQRLKYAOEE 818

#### RESULT 9

GNVYHM

genome polyprotein - human hepatitis A virus (strain HM-175, wild type)

N;Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core prot

B; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D

C;Species: human hepatitis A virus

A;Note: host Homo sapiens (man)

C;Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999

C;Accession: A25981

R;Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.

J. Virol. 61, 50-59, 1987

A;Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with dif

A;Reference number: A25981; MUID:87061253; PMID:3023706

A;Accession: A25981

A;Molecule type: genomic RNA

A;Residues: 1-2227 <COH>

A;Cross-references: EMBL:M14707; NID:G329582; PIDN:AAA45465.1; PID:G329583

C;Superfamily: hepatitis A virus genome polyprotein

C;Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltransf

F;1-23/Product: coat protein 1A #status predicted <VP4>

F;24-245/Product: coat protein 1B #status predicted <VP2>

F;246-491/Product: coat protein 1C #status predicted <VP3>

F;492-791/Product: coat protein 1D #status predicted <VP1>

F;792-980/Product: core protein 2A #status predicted <C2A>

F;981-1087/Product: core protein 2B #status predicted <C2B>

F;1088-1422/Product: core protein 2C #status predicted <C2C>

F;1423-1496/Product: protein 3A #status predicted <C3A>

F;1497-1519/Product: protein 3B #status predicted <C3B>

F;1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>

F;1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 102; DB 1; Length 2227;

Best Local Similarity 100.0%; Pred. No. 1.1e-07; Mismatches 20; Conservative 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAOEE 20

|||||

DB 799 PYKELRLEVGKQRLKYAOEE 818

#### RESULT 10

GNVYHR

genome polyprotein - human hepatitis A virus

N;Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core prot

NA polymerase (EC 2.7.7.48), protein 3D

C;Species: human hepatitis A virus

A;Note: host Homo sapiens (man)

C;Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999

C;Accession: A03903

R;Najarian, R.; Caput, D.; Ges, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van Nest,

Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985

A;Title: Primary structure and gene organization of human hepatitis A virus.

A;Reference number: A03903; MUID:85190549; PMID:2986127

A;Accession: A03903

A;Molecule type: genomic RNA

A;Residues: 1-2227 <NAJ>

A;Cross-references: GB:K02990; NID:G329596; PIDN:AAA45472.1; PID:G329597

C;Superfamily: hepatitis A virus genome polyprotein

C;Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltransf

F;1-245/Product: coat protein 1A #status predicted <C1A>

F;246-491/Product: coat protein 1B #status predicted <C1B>

F;492-836/Product: coat protein 1C #status predicted <C1C>

F;837-980/Product: core protein 2A #status predicted <C2A>

F;981-1076/Product: core protein 2B #status predicted <C2B>

F;1077-1422/Product: core protein 2C #status predicted <C2C>

F;1423-1484/Product: protein 3A #status predicted <C3A>

F;1485-1507/Product: protein 3B #status predicted <C3B>

```

F:24-246/Product: coat protein 1B #status predicted <VP2>
F:247-491/Product: coat protein 1C #status predicted <VP3>
F:432-836/Product: coat protein 1D #status predicted <VP1>
F:837-980/Product: coat protein 2A #status predicted <P2A>
F:981-1108/Product: core protein 2B #status predicted <P2B>
F:1109-1438/Product: core protein 2C #status predicted <P2C>
F:1439-1496/Product: protein 3A #status predicted <P3A>
F:1497-1519/Product: genome-linked protein VPg #status predicted <VPg>
F:1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>
F:1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match          100.0%; Score 102; DB 1; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
DB 799 PYKELRLEVGKQRLKYAQEE 818
|||||

RESULT 13
PQ0427
genome polypeptide - human hepatitis A virus (strain EP-35.730) (fragment)
C:Species: human hepatitis A virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: PQ0427
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel
J. Gen. Virol. 73, 1385-1377, 1992
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr
A:Reference number: PQ0427; PMID:92300330; PMID:1318940
A:Accession: PQ0427
A:Molecule type: mRNA
A:Residues: 156 <ROB>
C:Comment: This protein is from the VP1/2A Junction region.
C:Superfamily: hepatitis A virus genome polypeptide
C:Keywords: coat protein; core protein; polypeptide

Query Match          98.0%; Score 100; DB 2; Length 56;
Best Local Similarity 95.0%; Pred. No. 3.8e-09;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
DB 36 PYKELRLEVGKQRLKYAQEE 55
|||||

RESULT 14
PQ0429
genome polypeptide - human hepatitis A virus (strain PRC16) (fragment)
C:Species: human hepatitis A virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: PQ0429
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel
J. Gen. Virol. 73, 1385-1377, 1992
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr
A:Reference number: PQ0427; PMID:92300330; PMID:1318940
A:Accession: PQ0429
A:Molecule type: mRNA
A:Residues: 156 <ROB>
A:Note: this protein is from the VP1/2A Junction region
C:Superfamily: hepatitis A virus genome polypeptide
C:Keywords: coat protein; core protein; polypeptide

Query Match          97.1%; Score 99; DB 2; Length 56;
Best Local Similarity 95.0%; Pred. No. 5.5e-09;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
DB 36 PYKELRLEVGKQRLKYAQEE 55
|||||

RESULT 15

```



```

Query Match          93.1%; Score 95; DB 1; Length 2230;
Best Local Similarity 90.0%; Pred. No. 1.4e-06;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 PYKELRLVGVKQRLKYAQEE 20
         |||||:|||||:|||||
DB      803 PYKELMEVGVKQRLKYAMEE 822

RESULT 17
PQ0435
genome polypeptin - human hepatitis A virus (strain Cy145) (fragment)
C:Species: human hepatitis A virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: PQ0435
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel, J. Gen. Virol. 73, 1365-1377, 1992
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geographical regions
A:Reference number: PQ0427; MUID:92300330; PMID:1318940
A:Accession: PQ0435
A:Molecule type: mRNA
A:Residues: 1-55 <ROB>
A:Note: this protein is from the VP1/2A Junction region
C:Superfamily: hepatitis A virus genome polypeptin
C:Keywords: coat protein; core protein; polypeptin

Query Match          92.2%; Score 94; DB 2; Length 55;
Best Local Similarity 90.0%; Pred. No. 3.4e-08;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 PYKELRLVGVKQRLKYAQEE 20
         |||||:|||||:|||||
DB      35 PYKELRLVGVKQRLKYAMEE 54

RESULT 18
PQ0437
genome polypeptin - human hepatitis A virus (strain JM55) (fragment)
C:Species: human hepatitis A virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C:Accession: PQ0437
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel, J. Gen. Virol. 73, 1365-1377, 1992
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geographical regions
A:Reference number: PQ0427; MUID:92300330; PMID:1318940
A:Accession: PQ0437
A:Molecule type: mRNA
A:Residues: 1-56 <ROB>
A:Note: this protein is from the VP1/2A Junction region
C:Superfamily: hepatitis A virus genome polypeptin
C:Keywords: coat protein; core protein; polypeptin

Query Match          92.2%; Score 94; DB 2; Length 56;
Best Local Similarity 90.0%; Pred. No. 3.5e-08;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 PYKELRLVGVKQRLKYAQEE 20
         |||||:|||||:|||||
DB      36 PYKELRLVGVKQRLKYAMEE 55

RESULT 19
GNVY52
genome polypeptin - simian hepatitis A virus (strain CV-145) (fragment)
N:Contains: amino end of core protein 2A; coat protein 1A; coat protein 1B; coat protein 1C;
C:Species: simian hepatitis A virus
A:Note: host Macaca fascicularis (cynomolgus macaque)
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jul-1999
C:Accession: JQ1180
R:Nainan, O.V.; Margolis, H.S.; Robertson, B.H.; Balayan, M.; Brinton, M.A. J. Gen. Virol. 72, 1655-1689, 1991
A:Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus macaques

```

A:Reference number: JQ1180; MUID:913111421; PMID:1649902  
A:Accession: JQ1180  
A:Molecule type: Genomic RNA  
A:Residues: 1-839 <NAI>  
A:Cross-references: GB:M59286; NID:G329599; PID:AAA45473.1; PID:G555083  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; glycoprotein; polyprotein  
F:1-23/Product: coat protein 1A #status predicted <VP0>  
F:24-245/Product: coat protein 1B #status predicted <VP3>  
F:246-431/Product: coat protein 1C #status predicted <VP1>  
F:492-839/Product: core protein 2A (fragment) #status predicted <P2P>  
F:261,312,728,756/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 92.2%; Score 94; DB 1; Length 839;  
Best Local Similarity 90.0%; Pred. No. 6.9e-07;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PYKELRLEVGKQRLKYAQEE 20  
Db 798 PYKELRLEVGKQRLKYAQEE 817

RESULT 20  
PQ0431  
genome polyprotein - human hepatitis A virus (strain No. 4) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
A:Accession: PQ0431  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0431  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: This protein is from the VP1/2A junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 91.2%; Score 93; DB 2; Length 56;  
Best Local Similarity 95.0%; Pred. No. 5e-08;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PYKELRLEVGKQRLKYAQEE 20  
Db 36 PYKELRLEVGKQRLKYAQEE 55

RESULT 21  
JH0135  
genome polyprotein - human hepatitis A virus (strain MS-1) (fragment)  
N:Contains: amino end of core protein 2A; coat protein 1D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 08-Apr-1994  
A:Accession: JH0135  
R:Robertson, B.H.; Brown, V.K.; Bradley, D.W.  
Virus Res. 8, 309-316, 1987  
A:Title: Nucleic acid sequence of the VP1 region of attenuated MS-1 hepatitis A virus.  
A:Reference number: JH0135; MUID:88129044; PMID:2829458  
A:Accession: JH0135  
A:Molecule type: genomic RNA  
A:Residues: 1-319 <ROB>  
A:Cross-references: GB:M22821  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; glycoprotein; polyprotein  
F:1-300/Product: coat protein 1B #status predicted <CPD>  
F:301-319/Product: core protein 2A (fragment) #status predicted <C2A>  
F:237/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 60.8%; Score 62; DB 2; Length 319;  
Best Local Similarity 100.0%; Pred. No. 0.032;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PYKELRLEVGKQ 12  
Db 308 PYKELRLEVGKQ 319

RESULT 22  
T45997  
hypothetical protein F9D24.280 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 02-Sep-2000  
A:Accession: T45997  
R:D'Angelio, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Lemcke, K.;  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23011  
A:Accession: T45997  
A:Molecule type: DNA  
A:Status: preliminary  
A:Residues: 1-219 <DAN>  
A:Cross-references: EMBL:AL137081  
A:Experimental source: cultivar Columbia; BAC clone F9D24  
C:Genetics:  
A:Map position: 3  
A:Introns: 85/3  
A:Note: F9D24.280  
C:Superfamily: Arabidopsis thaliana hypothetical protein F9D24.210

Query Match 45.1%; Score 46; DB 2; Length 219;  
Best Local Similarity 50.0%; Pred. No. 7.9;  
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 3 KELRLEVGKQRLKYAQEE 20  
Db 167 KKKVETGKARLQRAEE 184

RESULT 23  
T23804  
hypothetical protein C06E4.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
A:Accession: T23804  
R:Du, Z.; Gattung, S.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid C06E4.  
A:Reference number: Z20688  
A:Accession: T23804  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-285 <DUZ>  
A:Cross-references: EMBL:U41277; PIDN:AAA82479.1; CESP:C06E4.8  
C:Genetics:  
A:Gene: CESP:C06E4.8  
A:Introns: 31/2; 127/3; 153/3; 212/3  
C:Superfamily: Caenorhabditis elegans hypothetical protein C06E4.8

Query Match 45.1%; Score 46; DB 2; Length 285;  
Best Local Similarity 60.0%; Pred. No. 11;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 4 ELRLEVGKQRLKYAQ 18  
Db 42 ELRRQIGKQKHVEQ 56

RESULT 24  
D72121  
ATP synthase chain D - Chlamydomonas reinhardtii (strain CWL029)  
C:Species: Chlamydomonas reinhardtii, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
A:Accession: D72121  
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A:Reference number: A72000; MUID:99206606; PMID:10192388  
 A:Accession: D72121  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-209 <ARN>  
 A:Cross-references: GB:AE001594; GB:AE001363; NID:94376341; PIDN:AA018243.1; PID:G437634  
 A:Experimental source: strain CWL029  
 C:Genetics:  
 A:Gene: atpD  
 C:Superfamily: H+-transporting ATPase chain D

Query Match 44.1%; Score 45; DB 2; Length 209;  
 Best Local Similarity 50.0%; Pred. No. 11;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 KLRLEVGKQRLKYAQEE 20

DB 133 KKVMAEVSKERLKILEE 150

RESULT 25

B86502  
 A:Title: ATP synthase subunit D [imported] - Chlamydia pneumoniae (strain J138)  
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Aug-2002  
 A:Accession: B86502  
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A:Reference number: A86491; MUID:20330349; PMID:10871362  
 A:Accession: B86502  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-209 <STO>  
 A:Cross-references: GB:BA000008; NID:98978463; PIDN:BA998300.1; GSPDB:GN00142  
 A:Experimental source: strain J138  
 C:Genetics:  
 A:Gene: atpD  
 C:Superfamily: H+-transporting ATPase chain D

Query Match 44.1%; Score 45; DB 2; Length 209;  
 Best Local Similarity 50.0%; Pred. No. 11;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 KLRLEVGKQRLKYAQEE 20

DB 133 KKVMAEVSKERLKILEE 150

RESULT 26

AI0819  
 A:Title: Probable exported protein STY2748 [imported] - Salmonella enterica subsp. enterica serov  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: This species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AI0819  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A:Accession: AI0819  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-179 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD02709.1; PID:gl6503724; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY2748

Query Match 43.1%; Score 44; DB 2; Length 179;

Best Local Similarity 60.0%; Pred. No. 13;  
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 6 RLEVGKQRLKYAQEE 20  
 DB 107 RLEIQKQLARAE 121

RESULT 27

C71701  
 A:Title: branched-chain amino acid aminotransferase (ilvE) RP428 - Rickettsia prowazekii  
 C:Species: Rickettsia prowazekii  
 C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
 R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.  
 Nature 396, 133-140, 1998  
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
 A:Reference number: A71630; MUID:99039499; PMID:9823893  
 A:Accession: C71701  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-290 <AND>  
 A:Cross-references: GB:AJ235271; GB:AJ235269; NID:G3868717; PIDN:CAA14885.1; PID:G386098;  
 A:Experimental source: strain Madrid E  
 C:Genetics:  
 A:Gene: ilvE; RP428  
 C:Superfamily: branched-chain-amino-acid aminotransferase

Query Match 43.1%; Score 44; DB 2; Length 290;  
 Best Local Similarity 61.1%; Pred. No. 23;  
 Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 KLRLEVGKQRLKYAQEE 20

DB 225 KSLCLEVSKERLKLAQIE 242

RESULT 28

T09938  
 A:Title: hypothetical protein T16L4.260 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 15-Oct-1999  
 C:Accession: T09938  
 R:Bavan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.  
 submitted to the Protein Sequence Database, June 1999  
 A:Reference number: Z16897  
 A:Accession: T09938  
 A:Molecule type: DNA  
 A:Residues: 1-776 <BEV>  
 A:Cross-references: EMBL:AL079344; GSPDB:GN00062; ATSP:T16L4.260  
 A:Experimental source: cultivar Columbia; BAC clone T16L4  
 C:Genetics:  
 A:Gene: ATSP:T16L4.260  
 A:Map position: 4  
 A:Introns: 308/3; 419/1; 462/3; 482/1; 574/3; 627/1; 724/3; 744/3

Query Match 43.1%; Score 44; DB 2; Length 776;  
 Best Local Similarity 52.9%; Pred. No. 67;  
 Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 4 ELRLVGKQRLKYAQEE 20

DB 569 ELRLYGKQLRRAERD 595

RESULT 29

H85609  
 A:Title: hypothetical protein Z1195 [imported] - Escherichia coli (strain O157:H7, substrain EDL95)  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A:Accession: AI0819  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-179 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD02709.1; PID:gl6503724; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY2748



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:24:26 ; Search time 6.75 Seconds  
(without alignment)  
154.282 Million cell updates/sec

Title: US-09-171-432A-40

Perfect score: 102

Sequence: 1 PYKELRLVGVQRLKYAQEE 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	102	100.0	341	1	POLG_HPAVL
2	102	100.0	808	1	POLG_HPAVC
3	102	100.0	852	1	POLG_HPAVC
4	102	100.0	2226	1	POLG_HPAV2
5	102	100.0	2226	1	POLG_HPAV4
6	102	100.0	2226	1	POLG_HPAV8
7	102	100.0	2227	1	POLG_HPAVH
8	102	100.0	2227	1	POLG_HPAVL
9	102	100.0	2227	1	POLG_HPAVM
10	95	93.1	2230	1	POLG_HPAVS
11	94	92.2	839	1	POLG_HPAVT
12	45	44.1	209	1	VATD_CHLPN
13	44	43.1	290	1	ILVE_RICPR
14	44	43.1	583	1	FOJO_DROME
15	44	43.1	720	1	CANC_MOUSE
16	43.5	42.6	1819	1	GCP6_HUMAN
17	43	42.2	208	1	UPP_AQUAE
18	43	42.2	480	1	DNAA_RHIME
19	43	42.2	3587	1	TYCB_BREPA
20	42.5	41.7	179	1	YFGI_ECOLI
21	42	41.2	290	1	ILVE_RICGN
22	42	41.2	752	1	PARC_ECOLI
23	42	41.2	752	1	PARC_SALTY
24	42	41.2	939	1	SVY_CHLTR
25	41	40.2	68	1	BGGS_HUMAN
26	41	40.2	68	1	BGGA_HUMAN
27	41	40.2	147	1	GST2_HUMAN
28	41	40.2	318	1	Y234_AQUAE
29	41	40.2	446	1	ATPD_MYCTU
30	40	39.2	69	1	MOP_HAEIN
31	40	39.2	213	1	PTH_STRPN
32	40	39.2	189	1	URK_YERPE
33	40	39.2	353	1	Y070_ARCFU

34	40	39.2	456	1	EF1A_DICDI
35	40	39.2	470	1	BFR1_YEAST
36	40	39.2	571	1	ILVD_STRMU
37	40	39.2	685	1	STM1_HUMAN
38	40	39.2	685	1	STM1_MOUSE
39	40	39.2	723	1	SYM_PTRAHO
40	40	39.2	733	1	ERG7_RAT
41	40	39.2	764	1	UBF1_HUMAN
42	40	39.2	764	1	UBF1_RAT
43	40	39.2	765	1	UBF1_MOUSE
44	40	39.2	795	1	LON_MYCPN
45	40	39.2	939	1	SYV_CHLMU
46	40	39.2	939	1	RPOA_LELV
47	39.5	38.7	976	1	SN21_HUMAN
48	39.5	38.7	1009	1	SN21_CAEEL
49	39.5	38.7	1812	1	BRC1_MOUSE
50	39	38.2	46	1	DIUH_LOCOMI
51	39	38.2	283	1	Y191_CIOAB
52	39	38.2	362	1	ILVE_STRCO
53	39	38.2	402	1	APL3_HUMAN
54	39	38.2	424	1	EXON_NEVOP
55	39	38.2	495	1	THIC_LEPIN
56	39	38.2	496	1	DNAA_BRUME
57	39	38.2	508	1	DNAA_BRUSU
58	39	38.2	508	1	V56K_PLRV1
59	39	38.2	508	1	V56K_PLRVW
60	39	38.2	553	1	PABP_SCHPO
61	39	38.2	1427	1	REST_HUMAN
62	39	38.2	1898	1	TRHY_HUMAN
63	39	38.2	1901	1	YCFL_TOBAC
64	38.5	37.7	416	1	RPSD_MICAE
65	38.5	37.7	641	1	DNAA_METSS
66	38	37.3	186	1	TNR2_ECOLI
67	38	37.3	226	1	PMT_MOUSE
68	38	37.3	239	1	PTH_RHIME
69	38	37.3	264	1	RPOD_SULAC
70	38	37.3	275	1	NADE_ECOLI
71	38	37.3	289	1	SGCD_MESAU
72	38	37.3	289	1	SGCD_MOUSE
73	38	37.3	290	1	SGCD_HUMAN
74	38	37.3	298	1	YN05_YEAST
75	38	37.3	365	1	RECF_CHLMU
76	38	37.3	406	1	YG43_YEAST
77	38	37.3	420	1	YAGA_SCHPO
78	38	37.3	480	1	118E_CUCMA
79	38	37.3	490	1	PIT_BUCAP
80	38	37.3	493	1	ALGE_PSESM
81	38	37.3	493	1	PIT_BUCAP
82	38	37.3	799	1	SCA_DROME
83	38	37.3	802	1	OPHI_MOUSE
84	38	37.3	1050	1	BUIB_HUMAN
85	38	37.3	1052	1	BUIB_MOUSE
86	38	37.3	1966	1	MYSB_CAEEL
87	37.5	36.8	495	1	TRME_TREPA
88	37.5	36.8	522	1	GAG_HVZG1
89	37.5	36.8	1139	1	NGAP_HUMAN
90	37	36.3	42	1	BGG7_MOUSE
91	37	36.3	68	1	BGG7_BOVIN
92	37	36.3	68	1	BGG7_HUMAN
93	37	36.3	69	1	BGG7_RAT
94	37	36.3	85	1	RS16_PSESM
95	37	36.3	103	1	RL23_AQUAE
96	37	36.3	118	1	RNPA_VIBPA
97	37	36.3	118	1	RNPA_VIBVU
98	37	36.3	122	1	YP48_METJA
99	37	36.3	140	1	Y517_METJA
100	37	36.3	172	1	YC46_PYRAB

ALIGNMENTS

RESULT 1

```

POLG_HPAV1          STANDARD;          PRT;    341 AA.
AC   P13672;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Genome polypeptide [Contains: Coat protein VP3 (1C); Coat protein VP1
DE   (1D); Core protein P2A] (Fragment).
OS   Hepatitis A virus (strain LDC-1).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
OX   NCBI_TaxID=12093;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89263805; PubMed=2542903;
RA   Andonov A.P., Lau P., Chaudhary R.;
RT   "Nucleotide sequence of the VP1 gene from a Chinese strain of
RT   hepatitis A virus (HAV).";
RL   Nucleic Acids Res. 17:3594-3594(1989).
CC   -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
CC   each of which is composed of one copy each of proteins VP1, VP2,
CC   VP3, and VP4.
CC   -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; X14666; CAA32794.1; -.
DR   PIR; S04137;
DR   InterPro; IPR008975; Viral_cap_coat.
KW   Polypeptide; Coat protein; Core protein.
FT   NON_TER 1
FT   CHAIN    <1 1 COAT PROTEIN VP3 (P1B).
FT   CHAIN    2 340 COAT PROTEIN VP1.
FT   CHAIN    341 >341 CORE PROTEIN P2A.
FT   NON_TER 341 341
FT   SEQUENCE 341 AA; 38003 MW; 066918289BF126D5 CRC64;
SQ
Query Match          100.0%; Score 102; DB 1; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy   1 PYKELRLEVGKQRLKYAQEE 20
Db   303 PYKELRLEVGKQRLKYAQEE 322

RESULT 2
POLG_HPAVG          STANDARD;          PRT;    808 AA.
AC   Q02381;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core protein
DE   P2A] (Fragment).
OS   Hepatitis A virus (strain GA76).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
OX   NCBI_TaxID=31706;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92260183; PubMed=1316423;
RA   Khanna B., Spelbring J.E., Innis B.L., Robertson B.H.;
RT   "Characterization of a genetic variant of human hepatitis A virus.";
RL   J. Med. Virol. 36:118-124(1992).
CC   -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
CC   each of which is composed of one copy each of proteins VP1, VP2,

```

```

CC   VP3, and VP4.
CC   -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; M66695; AAA45477.1; -.
DR   InterPro; IPR008975; Viral_cap_coat.
KW   Polypeptide; Coat protein; Core protein.
FT   NON_TER 1
FT   CHAIN    <1 2 COAT PROTEIN VP4 (P1A).
FT   CHAIN    3 223 COAT PROTEIN VP2 (P1B).
FT   CHAIN    224 470 COAT PROTEIN VP3 (P1C).
FT   CHAIN    471 770 COAT PROTEIN VP1 (P1D).
FT   CHAIN    771 >808 CORE PROTEIN P2A.
FT   NON_TER 808 808
FT   SEQUENCE 808 AA; 90632 MW; D80CE7E57A479C12 CRC64;
SQ
Query Match          100.0%; Score 102; DB 1; Length 808;
Best Local Similarity 100.0%; Pred. No. 2.6e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy   1 PYKELRLEVGKQRLKYAQEE 20
Db   778 PYKELRLEVGKQRLKYAQEE 797

RESULT 3
POLG_HPAVC          STANDARD;          PRT;    852 AA.
AC   P06442; Q83741; Q83742;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core protein
DE   P2A] (Fragment).
OS   Hepatitis A virus (strain CR326).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Hepatovirus.
OX   NCBI_TaxID=12097;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85185648; PubMed=2985793;
RA   Linemeyer D.L., Menke J.G., Martin-Gallardo A., Hughes J.V.,
RA   Young A., Mitra S.W.;
RT   "Molecular cloning and partial sequencing of hepatitis A viral cDNA.";
RL   J. Virol. 54:247-255(1985).
CC   -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
CC   each of which is composed of one copy each of proteins VP1, VP2,
CC   VP3, and VP4.
CC   -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC   -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; M10033; AAA45470.1; -.
DR   PIR; A03904; GNNYHA.
DR   InterPro; IPR008975; Viral_cap_coat.
KW   Polypeptide; Coat protein; Core protein.
FT   CHAIN    1 23 COAT PROTEIN VP4 (P1A).
FT   CHAIN    24 245 COAT PROTEIN VP2 (P1B).
FT   CHAIN    246 491 COAT PROTEIN VP3 (P1C).

```

FT CHAIN 492 836 COAT PROTEIN VP1 (P1D).  
 FT CHAIN 837 >852 CORE PROTEIN P2A.  
 FT NON\_TER 852 852  
 SQ SEQUENCE 852 AA; 95563 MW; 73D3ED0AD532820E CRC64;  
 Query Match 100.0%; Score 102; DB 1; Length 852;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
 DB 799 PYKELRLEVGKQRLKYAQEE 818

RESULT 4  
 POLG\_HPAV2 STANDARD; PRT: 2226 AA.  
 AC P26580;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
 DE P3D (EC 2.7.7.48).  
 OS Hepatitis A virus (strain 24a).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=12094;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91162758; PubMed=1705995;  
 RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
 RA Cromeans T., Jansen R.W.;  
 RA "Antigenic and genetic variation in cytopathic hepatitis A virus  
 RT variants arising during persistent infection: evidence for genetic  
 RT recombination.";  
 RL J. Virol. 65:2056-2065(1991).  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(N).  
 CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M59810; AAA45468.1; --  
 DR MEROPS: C03.005; --  
 DR InterPro: IPR004004; Calici\_pol\_hel.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR006005; RNA\_helicase.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR001205; RNA\_pol\_P3D.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR InterPro: IPR008975; Viral\_cap\_coat.  
 DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR Pfam: PF00910; RNA\_helicase; 1.  
 DR PRINTS: PR00918; CALICVIRUS.  
 DR Polyprotein; Coat protein; Core protein; Transferase;  
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
 FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
 FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
 FT CHAIN 245 491 COAT PROTEIN VP3 (P1C).  
 FT CHAIN 492 794 COAT PROTEIN VP1 (P1D).  
 FT CHAIN 795 900 CORE PROTEIN P2A.  
 FT CHAIN 901 1087 CORE PROTEIN P2B.

FT CHAIN 1088 1422 CORE PROTEIN P2C.  
 FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
 FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
 FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
 FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
 SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;  
 Query Match 100.0%; Score 102; DB 1; Length 2226;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
 DB 799 PYKELRLEVGKQRLKYAQEE 818

RESULT 5  
 POLG\_HPAV4 STANDARD; PRT: 2226 AA.  
 AC P26581;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
 DE P3D (EC 2.7.7.48).  
 OS Hepatitis A virus (strain 43c).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=12095;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91162758; PubMed=1705995;  
 RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
 RA Cromeans T., Jansen R.W.;  
 RA "Antigenic and genetic variation in cytopathic hepatitis A virus  
 RT variants arising during persistent infection: evidence for genetic  
 RT recombination.";  
 RL J. Virol. 65:2056-2065(1991).  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(N).  
 CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M59809; AAA45469.1; --  
 DR MEROPS: C03.005; --  
 DR InterPro: IPR004004; Calici\_pol\_hel.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR006005; RNA\_helicase.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR001205; RNA\_pol\_P3D.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR InterPro: IPR008975; Viral\_cap\_coat.  
 DR Pfam: PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR Pfam: PF00910; RNA\_helicase; 1.  
 DR PRINTS: PR00918; CALICVIRUS.  
 DR Polyprotein; Coat protein; Core protein; Transferase;  
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
 FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
 FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
 FT CHAIN 245 491 COAT PROTEIN VP3 (P1C).  
 FT CHAIN 492 794 COAT PROTEIN VP1 (P1D).  
 FT CHAIN 795 900 CORE PROTEIN P2A.  
 FT CHAIN 901 1087 CORE PROTEIN P2B.

FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA9B09BF75 CRC64;  
Query Match 100.0%; Score 102; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 7.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 PYKELRLEVKGQRLKYAQEE 20  
Db 799 PYKELRLEVKGQRLKYAQEE 818  
RESULT 6  
POLG HPAV8 STANDARD; PRT; 2226 AA.  
ID POLG HPAV8 STANDARD; PRT; 2226 AA.  
AC P26582;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
DE P3D (EC 2.7.7.48).  
OS Hepatitis A virus (strain 18f).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12096;  
RN [1]  
PP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
RA Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination.";  
RL J. Virol. 65:2056-2065(1991).  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC [RNA] (N).  
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M59808; AAA45467.1; -;  
CC PDB; 1O47; 15-MAY-00.  
CC MEQOPS; C03.005; -;  
CC InterPro; IPR004004; Calici pol hel.  
CC InterPro; IPR009003; Cys ser trypsin.  
CC InterPro; IPR000605; RNA helicase.  
CC InterPro; IPR007095; RNA pol DS PS.  
CC InterPro; IPR001205; RNA pol P3D.  
CC InterPro; IPR007094; RNA pol-psvir.  
CC InterPro; IPR008975; Viral cap coat.  
CC Pfam; PF00680; RNA dep RNA pol; 1.  
CC Pfam; PF00910; RNA helicase; 1.  
CC PRINTS; PR00918; CALICIVIRUS.  
KW Polyprotein; Coat protein; Core protein; Transferrase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.  
FT CHAIN 1 23 COAT PROTEIN VP4 (PIA).

FT CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
FT CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
FT CHAIN 492 794 COAT PROTEIN VP1 (PID).  
FT CHAIN 795 900 CORE PROTEIN P2A.  
FT CHAIN 901 1087 CORE PROTEIN P2B.  
FT CHAIN 1088 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2226 AA; 251292 MW; 24964A6396C8D6B CRC64;  
Query Match 100.0%; Score 102; DB 1; Length 2226;  
Best Local Similarity 100.0%; Pred. No. 7.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 PYKELRLEVKGQRLKYAQEE 20  
Db 799 PYKELRLEVKGQRLKYAQEE 818  
RESULT 7  
POLG HPAVH STANDARD; PRT; 2227 AA.  
ID POLG HPAVH STANDARD; PRT; 2227 AA.  
AC P08617; P06443; Q81082;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
DE P3D (EC 2.7.7.48).  
OS Hepatitis A virus (strain HM-175).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12098;  
RN [1]  
PP SEQUENCE FROM N.A.  
RX STRAIN=Wild type;  
RX MEDLINE=87061253; PubMed=3023706;  
RA Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A.,  
RA Baroudy B.M.;  
RT "Complete nucleotide sequence of wild-type hepatitis A virus:  
RT comparison with different strains of hepatitis A virus and other  
RT picornaviruses.";  
RL J. Virol. 61:50-59(1987).  
RN [2]  
PP SEQUENCE FROM N.A.  
RX STRAIN=Attenuated;  
RX MEDLINE=87175701; PubMed=3031686;  
RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstone S.M.,  
RA Purcell R.H.;  
RT "Complete nucleotide sequence of an attenuated hepatitis A virus:  
RT comparison with wild-type virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501(1987).  
RN [3]  
PP SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.  
RX MEDLINE=85166289; PubMed=2984684;  
RA Baroudy B.M., Ticehurst J.R., Miele T.A., Maizel J.V. Jr.,  
RA Purcell R.H., Feinstone S.M.;  
RT "Sequence analysis of hepatitis A virus cDNA coding for capsid  
RT proteins and RNA polymerase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC [RNA] (N).  
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -1- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED  
CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -1- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT  
CC SHOWN.



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M14114; AAA45475.1; -;  
 CC EMBL; M14707; AAA45465.1; -;  
 CC EMBL; M14707; AAA45466.1; ALT\_INIT.  
 CC EMBL; M16632; AAA45471.1; -;  
 CC PIR; A03905; A03905.  
 CC PIR; A25981; GNNYHM.  
 CC PIR; A94149; GNNYMK.  
 CC PDB; 1HAV; 23-DEC-96.  
 CC MEROPS; C03.005; -;  
 CC InterPro; IPR004004; Calici\_pol\_hel.  
 CC InterPro; IPR009003; Cys\_Ser\_trypsin.  
 CC InterPro; IPR006005; RNA\_helicase.  
 CC InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 CC InterPro; IPR001205; RNA\_pol\_P3D.  
 CC InterPro; IPR007094; RNA\_pol\_PSVir.  
 CC InterPro; IPR008975; Viral\_cap\_coat.  
 CC Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 CC Pfam; PF00910; RNA\_helicase; 1.  
 CC PRINTS; PR00918; CALICIVIRUS.  
 CC Polyprotein; Coat protein; Core protein; Transferase;  
 CC RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.  
 CC CHAIN 1 23  
 CC CHAIN 24 245  
 CC CHAIN 246 491  
 CC CHAIN 492 836  
 CC CHAIN 837 980  
 CC CHAIN 981 1087  
 CC CHAIN 1088 1422  
 CC CHAIN 1423 1496  
 CC CHAIN 1497 1519  
 CC CHAIN 1520 1738  
 CC CHAIN 1739 2227  
 CC CHAIN 77 77  
 CC CHAIN 764 764  
 CC CHAIN 821 821  
 CC CHAIN 1052 1052  
 CC CHAIN 1062 1062  
 CC CHAIN 1118 1118  
 CC CHAIN 1163 1163  
 CC CHAIN 1277 1277  
 CC CHAIN 1500 1500  
 CC CHAIN 1805 1805  
 CC CHAIN 1930 1930  
 CC CHAIN 2227 2227  
 CC CHAIN 251506 MW; 01E225E7ABEB740A6 CRC64;

Query Match 100.0%; Score 102; DB 1; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVKGKRLKYAQEE 20  
 |||||  
 Db 799 PYKELRLEVKGKRLKYAQEE 818

## RESULT 8

ID POLG\_HPAVL STANDARD; PRT; 2227 AA.  
 AC P06441;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [contains: Coat proteins VP1 to VP4; Core proteins  
 DE P2A to P2C; Probable proteins P3A to P3C; RNA-directed RNA polymerase

DE P3D (EC 2.7.7.48)].  
 OS Hepatitis A virus (strain LA).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_taxID=12099;  
 RP [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=85190549; PubMed=2986127;  
 RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,  
 RA Merryweather J., van Nest G., Dina D.;  
 RT "Primary structure and gene organization of human hepatitis A virus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(N).  
 CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; K02990; AAA45472.1; -;  
 CC PIR; A03903; GNNYHR.  
 CC MEROPS; C03.005; -;  
 CC InterPro; IPR004004; Calici\_pol\_hel.  
 CC InterPro; IPR009003; Cys\_Ser\_trypsin.  
 CC InterPro; IPR006005; RNA\_helicase.  
 CC InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 CC InterPro; IPR001205; RNA\_pol\_P3D.  
 CC InterPro; IPR007094; RNA\_pol\_PSVir.  
 CC InterPro; IPR008975; Viral\_cap\_coat.  
 CC Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 CC Pfam; PF00910; RNA\_helicase; 1.  
 CC PRINTS; PR00918; CALICIVIRUS.  
 CC Polyprotein; Coat protein; Core protein; Transferase;  
 CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
 CC CHAIN 1 23  
 CC CHAIN 24 245  
 CC CHAIN 246 491  
 CC CHAIN 492 836  
 CC CHAIN 837 980  
 CC CHAIN 981 1076  
 CC CHAIN 1077 1422  
 CC CHAIN 1423 1484  
 CC CHAIN 1485 1507  
 CC CHAIN 1508 1678  
 CC CHAIN 1679 2227  
 CC CHAIN 2227 2227  
 CC CHAIN 251898 MW; 99A7354B4CD2799C CRC64;

Query Match 100.0%; Score 102; DB 1; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVKGKRLKYAQEE 20  
 |||||  
 Db 799 PYKELRLEVKGKRLKYAQEE 818

## RESULT 9

ID POLG\_HPAVM STANDARD; PRT; 2227 AA.  
 AC P13901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;  
 AC Q81090; Q81091; Q81092; Q81093;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
DE Hepatitis A virus (strain MBB).  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12100;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88045071; PubMed=2823500;  
RA Paul A.V., Tada H., der Helm K., Wissel T., Kiehn R., Wimmer E.,  
RA Deinhardt F.;  
RT "The entire nucleotide sequence of the genome of human hepatitis A  
RT virus (isolate MBB).";  
RL Virus Res. 8:153-171(1987).  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC {RNA}(N).  
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT  
CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M20273; AAA45474.1; -;  
DR MEROPS; C03.005; -;  
DR InterPro; IPR004004; Calici\_pol\_hel.  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR006005; RNA\_helicase.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR007094; RNA\_pol\_P3vir.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICIVIRUSNS.  
DR PolyProtein; Coat protein; Core protein; Transferase;  
DR RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1496  
FT CHAIN 1497 1519  
FT CHAIN 1520 1738  
FT CHAIN 1739 2227  
SQ SEQUENCE 2227 AA; 251425 MW; EC983ED2A7C86349 CRC64;  
  
Query Match 100.0%; Score 102; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 7.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
DQ 799 PYKELRLEVGKQRLKYAQEE 818  
  
RESULT 10  
POLG HPAAVS STANDARD; PRT; 2230 AA.  
ID POLG HPAAVS  
AC P14553;

DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
DE P3D (EC 2.7.7.48)].  
DE Simian hepatitis A virus (strain AGM-27).  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12102;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91311420; PubMed=1649901;  
RA Tsarev S.A., Emerson S.U., Balayan M.S., Ticehurst J.R.,  
RA Purcell R.H.;  
RT "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome  
RT structure and growth in cell culture with other HAV strains.";  
RL J. Gen. Virol. 72:1677-1683(1991).  
RN [2]  
RP SEQUENCE OF 1750-2164 FROM N.A.  
RX MEDLINE=89232168; PubMed=2541023;  
RA Balayan M.S., Kusov Y.Y., Andjapadze A.G., Tsarev S.A.,  
RA Sverdlov B.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;  
RT "Variations in genome fragments coding for RNA polymerase in human  
RT and simian hepatitis A viruses.";  
RL FEBS Lett. 247:425-428(1989).  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC {RNA}(N).  
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D00924; BAA00766.1; -;  
DR EMBL; X15461; GAA33490.1; -;  
DR PIR; A30470; GNNYSA.  
DR MEROPS; C03.005; -;  
DR InterPro; IPR004004; Calici\_pol\_hel.  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR006005; RNA\_helicase.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR007094; RNA\_pol\_P3vir.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICIVIRUSNS.  
DR PolyProtein; Coat protein; Core protein; Transferase;  
DR RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 27  
FT CHAIN 28 249  
FT CHAIN 250 495  
FT CHAIN 496 795  
FT CHAIN 796 984  
FT CHAIN 985 1091  
FT CHAIN 1092 1426  
FT CHAIN 1427 1498  
FT CHAIN 1499 1521  
FT CHAIN 1522 1741  
FT CHAIN 1742 2230  
SQ SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F19 CRC64;  
  
Query Match 93.1%; Score 95; DB 1; Length 2230;  
Best Local Similarity 90.0%; Pred. No. 9.3e-07;

Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKORLYAQEE 20  
 DB 803 PYKELRLEVGKORLYAQEE 822

RESULT 11

POLG\_HPAVT  
 ID POLG\_HPAVT STANDARD; PRT; 839 AA.

AC P31788;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein P2A] (Fragment).  
 OS Simian hepatitis A virus (strain CY-145).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OC NCBI\_TaxID=31707;  
 RN [1]

SEQUENCE FROM N.A.  
 RC MEDLINE=91311421; PubMed=1649902;  
 RA Nainan O.V., Margolis H.S., Robertson B.H., Balayan M., Brinton M.A.;  
 RT "Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus macaques (Macaca fascicularis).";  
 RL J. Gen. Virol. 72:1685-1689(1991).  
 CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units, each of which is composed of one copy each of proteins VP1, VP2, VP3, and VP4.  
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M59286; AAA45473.1; -;  
 CC PIR; JQ1180; GNNYS2.  
 CC InterPro; IPR008975; Viral\_cap\_coat.  
 KW Polyprotein; Coat protein; Core protein.  
 FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
 FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
 FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
 FT CHAIN 492 ? COAT PROTEIN VP1 (P1D).  
 FT CHAIN ? >839 CORE PROTEIN P2A.  
 FT NON TER 839 839  
 CC SEQUENCE 839 AA; 93825 MW; 2CACC4BD1E192DBC CRC64;

Query Match 92.2%; Score 94; DB 1; Length 839;  
 Best Local Similarity 90.0%; Pred. No. 4.9e-07;  
 Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKORLYAQEE 20  
 DB 798 PYKELRLEVGKORLYAQEE 817

RESULT 12

VATD\_CHLEN  
 ID VATD\_CHLEN STANDARD; PRT; 209 AA.

AC Q92991;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE V-type ATP synthase subunit D (EC 3.6.3.14) (V-type ATPase subunit D).  
 GN ATPD OR CPN090 OR CP0684 OR CPB090.  
 OS Chlamydia pneumoniae (Chlamydiales; Chlamydiales).  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiales; Chlamydiales.  
 OC NCBI\_TaxID=83558;

SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]

SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]

SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [4]

SEQUENCE FROM N.A.  
 RC STRAIN=TW-183;  
 RA Geng M.M., Schumacher A., Muehlendorfer I., Bensch K.W., Schaefer K.P.,  
 RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;  
 RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with other Chlamydia strains based on whole genome sequence analysis.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Produces ATP from ADP in the presence of a proton gradient across the membrane.  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate + H(+) (Out).  
 CC -1- SIMILARITY: Belongs to the V-ATPase D subunit family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AE001594; AAD18243.1; -;  
 CC EMBL; AE002227; AAF73696.1; -;  
 CC EMBL; AP002545; BAA98300.1; -;  
 CC EMBL; AE017157; AAF98023.1; ALT\_INIT.  
 CC PIR; B86502; B86502.  
 CC PIR; D72121; D72121.  
 CC TIGR; CP0684; -;  
 CC HAMAP; MF\_00271; -; 1.  
 CC InterPro; IPR002699; ATPsynth\_Dsub.  
 CC Pfam; PF01813; ATP-synt\_D; 1.  
 CC TIGRFAMs; TIGR00309; V\_Atpase\_subD; 1.  
 KW Hydrolase; ATP synthesis; Hydrogen ion transport; Complete proteome.  
 CC SEQUENCE 209 AA; 23788 MW; D40A5C36313FEEBF CRC64;

Query Match 44.1%; Score 45; DB 1; Length 209;  
 Best Local Similarity 50.0%; Pred. No. 6.5;  
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 KKLRLVGVKORLYAQEE 20  
 DB 133 KKYNAEVSERLKLLEE 150

## RESULT 13

ILVE\_RICPR STANDARD; PRT; 290 AA.

AC 005970; Q24176; PRT; 583 AA.

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable branched-chain amino acid aminotransferase (EC 2.6.1.42) (BCAT).

GN ILVE OR RP428.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI\_TaxID=782;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Madrid E;

RX MEDLINE=97419517; PubMed=9274032;

RA Andersson J.O., Andersson S.G.B.;

RT "Genomic rearrangements during evolution of the obligate intracellular parasite Rickettsia prowazekii as inferred from an analysis of 52015 bp nucleotide sequence.";

RT Microbiology 143:2783-2795 (1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Madrid E;

RX MEDLINE=99039499; PubMed=9823893;

RA Andersson S.G.B., Zomorodipour A., Andersson J.O.,

RA Sierhartz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of mitochondria.";

RT Nature 396:133-140 (1998).

CC -!- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).

CC -!- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-oxopentanoate + L-glutamate.

CC -!- COFACTOR: Pyridoxal phosphate.

CC -!- PATHWAY: Valine and isoleucine biosynthesis.

CC -!- SIMILARITY: Belongs to class-IV of pyridoxal-phosphate-dependent aminotransferases.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; Y11777; CAA72450.1; -.

DR EMBL; AJ235271; CAA14885.1; -.

DR PIR; C71701; C71701.

DR HSP; P00510; IA3G.

DR InterPro; IPR001544; Aminotrans\_IV.

DR Pfam; PF01063; aminotran\_4; 1.

DR ProDom; PD001961; Aminotran\_4; 1.

DR PROSITE; PS00770; AA\_TRANSF\_4; 1.

KW Transferase; Aminotransferase; Branched-chain amino acid biosynthesis; Pyridoxal phosphate; Complete proteome.

FT BINDING 155 155 PYRIDOXAL PHOSPHATE (BY SIMILARITY).

SQ SEQUENCE 290 AA; 33068 MW; 98374E325350763D CRC64;

Query Match 43.1%; Score 44; DB 1; Length 290;

Best Local Similarity 61.1%; Pred. No. 13;

Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 KELLREVGKQRLKYAQE 20

DB 225 KSLCLEVKEERLKLQAIE 242

## RESULT 14

## FOJO DROME

ID FOJO DROME STANDARD; PRT; 583 AA.

AC P54360; Q24176; PRT; 583 AA.

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Four-jointed protein.

GN FJ.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Eye imaginal disk;

RX MEDLINE=96038089; PubMed=7555705;

RA Villano J.L., Katz F.N.;

RT "Four-jointed is required for intermediate growth in the proximal-distal axis in Drosophila.";

RL Development 121:2767-2777 (1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Canton-S;

RX MEDLINE=96187865; PubMed=8606003;

RA Brodsky M.H., Steller H.;

RT Positional information along the dorsal-ventral axis of the Drosophila eye: Graded expression of the four-jointed gene.";

RL Dev. Biol. 173:428-446 (1996).

CC -!- FUNCTION: Required for intermediate growth in the proximal-distal axis. May be required for cell-cell signaling during disk development.

CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; U28837; AAA69524.1; -.

DR EMBL; U44904; BAB01809.1; -.

DR FlyBase; FBgn0000658; fj.

DR GO; GO:0005886; C:plasma membrane; IDA.

DR GO; GO:0007267; P:cell-cell signaling; IDA.

DR GO; GO:0045198; P:establishment of epithelial cell polarity; IMP.

DR GO; GO:0007446; P:imaginal disc growth; IMP.

DR GO; GO:0016348; P:leg joint morphogenesis (sensu Holometabola); IMP.

DR GO; GO:0007474; P:wing vein specification; IMP.

DR GO; GO:0007474; P:wing vein specification; IMP.

KW Transmembrane; Signal-anchor.

FT DOMAIN 1 78 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 79 99 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).

FT FT EXTRACELLULAR (POTENTIAL).

FT CARBOHYD 100 583 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 310 310 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 379 379 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 491 491 T -> S (IN REF. 2).

FT CONFLICT 126 126 M -> L (IN REF. 2).

FT CONFLICT 193 193 P -> R (IN REF. 2).

FT CONFLICT 288 288 R -> A (IN REF. 2).

FT CONFLICT 330 330 R -> A (IN REF. 2).

SQ SEQUENCE 583 AA; 65504 MW; 82F1EA2A299DB284 CRC64;

Query Match 43.1%; Score 44; DB 1; Length 583;

Best Local Similarity 50.0%; Pred. No. 27;

Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 YKELREVGKQRLKYAQE 19

DB 2 YDIKRLFEAGQOKLQQAQ 19

EMBL: AJ289241; CAC10067.1; --  
 EMBL: AJ289241; CAC10068.1; --  
 EMBL: AJ289243; CAC10070.1; --  
 EMBL: BC028751; AAH28751.1; --  
 HSSP: P04574; 1ALV.  
 MEROPS: C02.017; --  
 MGI: 1891369; Capn12.  
 InterPro: IPR002048; EF-hand.  
 InterPro: IPR001300; Peptidase\_C2.  
 InterPro: IPR00169; SHprot\_acsite.  
 Pfam: PF01067; Calpain\_III; 1.  
 Pfam: PF00648; Peptidase\_C2; 1.  
 PRINTS: PR00704; CALPAIN\_III; 1.  
 SMART: SM00720; calpain\_III; 1.  
 SMART: SM00230; CysPC; 1.  
 PROSITE: PS0203; CALPAIN\_CAT; 1.  
 PROSITE: PS00018; EF\_HAND; 1.  
 PROSITE: PS00640; THIOL\_PROTEASE\_ASN; FALSE\_NEG.  
 PROSITE: PS00139; THIOL\_PROTEASE\_CVS; 1.  
 PROSITE: PS00639; THIOL\_PROTEASE\_HIS; FALSE\_NEG.  
 Hydrolase; Thiol protease; Calcium-binding; Multigene family;  
 Alternative splicing.  
 DOMAIN 45 341 CALPAIN CATALYTIC.  
 DOMAIN 342 541 DOMAIN III.  
 DOMAIN 542 720 DOMAIN IV.  
 CA\_BIND 634 645 EF-HAND (POTENTIAL).  
 ACT\_SITE 105 105 BY SIMILARITY.  
 ACT\_SITE 259 259 BY SIMILARITY.  
 ACT\_SITE 283 283 BY SIMILARITY.  
 VARSPLIC 341 448  
 VARSPLIC 449 720  
 VARSPLIC 460 462  
 VARSPLIC 463 720  
 VARSPLIC 460 502  
 VARSPLIC 503 720  
 SEQUENCE 720 AA; 80588 MW; 37C07BDF0145B531 CRC64;  
 Query Match 43.1%; Score 44; DB 1; Length 720;  
 Best Local Similarity 50.0%; Pred. No. 33;  
 Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 PYKELRLEVQKQLKYAQQE 20  
 Db 545 PYKLELELAQLFELAGEE 564  
 RESULT 16  
 GCP6\_HUMAN  
 ID GCP6\_HUMAN STANDARD; PRT; 1819 AA.  
 AC Q9ERT7; Q9BY91; Q9UGX3; Q9UGX4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Gamma-tubulin complex component 6 (GCP-6).  
 GN TUBGCP6 OR GCP6 OR KIAA1669.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]

RESULT 15  
 CANC\_MOUSE  
 ID CANC\_MOUSE STANDARD; PRT; 720 AA.  
 AC Q9ER56; Q9ER53; Q9ER54; Q9ER55;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Calpain 12 (EC 3.4.22.-).  
 GN CAPN12.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).  
 MEDLINE=20422668; PubMed=10964513;  
 RA Dear T.N., Meier N.T., Hunn M., Boehm T.;  
 RT "Gene structure, chromosomal localization and expression pattern of  
 Capn12, a new member of the calpain large subunit gene family";  
 RL Genomics 68:152-160(2000).  
 RN [2]  
 SEQUENCE FROM N.A. (ISOFORM 2).  
 TISSUE=Mammary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Klausner R.L., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Bustow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Mullahy S.J.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramo R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.N.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease (By  
 similarity).  
 CC -!- CATALYTIC ACTIVITY: Broad endopeptidase specificity.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Name=1;  
 CC IsoId=Q9ER56-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9ER56-2; Sequence=VSP\_007807, VSP\_007808;  
 CC Name=3; Synonyms=variant 3;  
 CC IsoId=Q9ER56-3; Sequence=VSP\_007809, VSP\_007810;  
 CC Name=4; Synonyms=variant 2;  
 CC IsoId=Q9ER56-4; Sequence=VSP\_007811, VSP\_007812;  
 CC -!- TISSUE SPECIFICITY: Expression localized to the cortex of the hair  
 follicle during the anagen phase of hair cycle.  
 CC -!- SIMILARITY: Belongs to peptidase family C2.  
 CC -!- SIMILARITY: Contains 1 EF-hand calcium-binding domain.  
 CC -!- SIMILARITY: Contains 1 Calpain catalytic domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL: AJ289241; CAC10066.1; --

RP SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.  
RX MEDLINE=2151508; PubMed=11694571;  
RA Murphy S.M., Preble A.M., Patel U.K., O'Connell K.L., Dias D.P.,  
RA Moritz M., Agard D., Stults J.T., Stearns T.;  
RT "GCP5 and GCP6: two new members of the human gamma-tubulin complex.";  
RL Mol. Biol. Cell 12:3340-3352(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20057165; PubMed=10591208;  
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,  
RA Clamp M., Smink L.J., Alnsough R., Almeida J.P., Babbage A.K.,  
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beagley O.P.,  
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,  
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,  
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,  
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,  
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
RA Soderlund C., Spragon L., Steward C.A., Sultana J.E., Swann R.M.,  
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
RA Shinohara A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
RA Zhan M., Zhang G., Chiswick S., Murray J., Miller N., Minx P.,  
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
RA Hinds K., Kemp K., Latreille P., Layman D., Oersky P., Rohlfing T.,  
RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,  
RA Korf I., Bedell J.A., Hillier L.M., Mardis E., Waterston R.,  
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,  
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanaki J.P.,  
RA Peyrard K., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,  
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
RA Khan A.S., Lane L., Tilahun Y., Wright H.;  
RT "The DNA sequence of human chromosome 22.";  
RL Nature 402:489-495(1999).  
RN [3]  
RN SEQUENCE OF 33-1819 FROM N.A. (ISOFORM 2).  
RP TISSUE=Brain;  
RX MEDLINE=21156230; PubMed=11258795;  
RA Hirosewa M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;  
RT "Identification of novel transcribed sequences on human chromosome 22  
by expressed sequence tag mapping.";  
RL DNA Res. 8:1-9(2001).  
CC -!- FUNCTION: Gamma-tubulin complex is necessary for microtubule  
nucleation at the centrosome.  
CC -!- SUBUNIT: Gamma-tubulin complex is composed of gamma-tubulin, GCP2,  
GCP3, GCP4, GCP5 and GCP6.  
CC -!- SUBCELLULAR LOCATION: Centrosome.  
CC -!- ALTERNATIVE PRODUCTS:  
Event=alternative splicing, Named isoforms=2;  
CC Name=1;  
CC IsoId=Q96R77-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q96R77-2; Sequence=VSP\_001624;  
CC Notes=No experimental confirmation available;  
CC -!- SIMILARITY: Belongs to the GCP family.

CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts  
in positions 1371 and 1758.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AF272887; AAK92968.1; -;  
DR EMBL; AL022326; CAB83046.1; ALT\_SEQ.  
DR EMBL; AL022328; CAB83047.1; ALT\_SEQ.  
DR EMBL; AB051456; BAB33339.1; ALT\_FRAME.  
DR Genbank; HGNC:18127; TUBGCP6.  
DR GO; GO:0008274; C:gamma-tubulin ring complex; IDA.  
DR GO; GO:0008017; F:microtubule binding; IDA.  
DR GO; GO:0007020; P:microtubule nucleation; IDA.  
DR InterPro; IPR007259; Spc97\_Spc98.  
DR Pfam; PF04130; Spc97\_Spc98; 1.  
KW Microtubule; Repeat; Alternative splicing.  
FT DOMAIN 1027 1269 9 X 27 AA TANDEM REPEATS.  
FT REPEAT 1027 1053 1.  
FT REPEAT 1054 1080 2.  
FT REPEAT 1081 1107 3.  
FT REPEAT 1108 1134 4.  
FT REPEAT 1135 1161 5.  
FT REPEAT 1162 1188 6.  
FT REPEAT 1189 1215 7.  
FT REPEAT 1216 1242 8.  
FT REPEAT 1243 1269 9.  
FT VARSPLIC 1724 1757 Missing (in isoform 2).  
FT CONFLICT 567 567 S -> L (IN REF. 3).  
FT CONFLICT 1377 1377 A -> T (IN REF. 2 AND 3).  
FT CONFLICT 1621 1621 L -> V (IN REF. 2).  
SQ SEQUENCE 1819 AA; 200455 MW; 923576544D34594A CRC64;  
Query Match 42.6%; Score 43.5; DB 1; Length 1819;  
Best Local Similarity 55.6%; Pred. No. 1e+02;  
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;  
QY 3 KELRLVGVKQRL-KYAOE 19  
DB 661 KELRLVGVKQRL-KYAOE 678  
RESULT 17  
UPP\_AQUAE  
ID UPP\_AQUAE STANDARD; PRT; 208 AA.  
AC O67914;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Uracil phosphoribosyltransferase (EC 2.4.2.9) (UMP pyrophosphorylase)  
DE (UPRTase).  
GN UPP OR URAP OR AQ\_2163.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VFS;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Snead M.A., Kaller M., Aujay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
aeolicus.";  
RL Nature 392:353-358(1998).  
CC -!- CATALYTIC ACTIVITY: UMP + diphosphate = uracil + 5-phospho-  
alpha-D-ribose 1-diphosphate.



```

CC DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-
CC ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC -!- SIMILARITY: Contains 3 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004835; AAC45929.1; -
CC HSSP; P14687; 1AMU.
CC InterPro; IPR000873; AMP-bind.
CC InterPro; IPR001242; Condensatn.
CC InterPro; IPR006163; Pp_bind.
CC InterPro; IPR006162; Pplantn S.
CC Pfam; PF00501; AMP-binding; 3.
CC Pfam; PF00668; Condensation; 4.
CC Pfam; PF00550; PP-binding; 3.
CC PRINTS; PR00154; AMPBINDING.
CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
CC PROSITE; PS00455; AMP BINDING; 3.
CC PROSITE; PS50075; ACP DOMAIN; 3.
CC Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
CC Multifunctional enzyme; Repeat.
CC REPEAT 466 1045 DOMAIN 1 (PROLINE-ACTIVATING).
CC REPEAT 1522 2081 DOMAIN 2 (PHENYLALANINE-ACTIVATING).
CC REPEAT 2540 3122 DOMAIN 3 (D-PHENYLALANINE-ACTIVATING).
CC DOMAIN 977 1044 ACYL CARRIER (ACP) 1.
CC DOMAIN 2012 2079 ACYL CARRIER (ACP) 2.
CC DOMAIN 3045 3111 ACYL CARRIER (ACP) 3.
CC BINDING 1007 1007 PHOSPHOPANTETHEINE (BY SIMILARITY).
CC BINDING 2042 2042 PHOSPHOPANTETHEINE (BY SIMILARITY).
CC BINDING 3075 3075 PHOSPHOPANTETHEINE (BY SIMILARITY).
CC SEQUENCE 3587 AA; 404810 MW; 322B8471BBB28B47 CRC64;
Query Match 42.2%; Score 43; DB 1; Length 3587;
Best Local Similarity 50.0%; Pred. No. 2.5e+02;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 4 ELRLEVKQRLKYAQE 19
DB 3522 ELRISGVSRQLQVEK 3537
RESULT 20
YFGI_ECOLI STANDARD; PRT; 179 AA.
AC P76573;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yfgi.
GN YFGI OR B2506.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";

```

```

RL Science 277:1453-1474 (1997).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000337; AAC75559.1; -
CC PIR; A65027; A65027.
CC Ecogene; EG14205; yfgi.
CC KW Hypothetical protein; Complete proteome.
CC SEQUENCE 179 AA; 20230 MW; ABB6FE42D5BEB283 CRC64;
Query Match 41.7%; Score 42.5; DB 1; Length 179;
Best Local Similarity 61.1%; Pred. No. 14;
Matches 11; Conservative 3; Mismatches 3; Indels 1; Gaps 1;
QY 4 ELR-LEVGKQRLKYAQEE 20
DB 104 ELRSLRIQKQKLAKKE 121
RESULT 21
ILVE_RICCN STANDARD; PRT; 230 AA.
ID ILVE_RICCN
AC Q92126;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable branched-chain amino acid aminotransferase (EC 2.6.1.42)
DE (BCAT).
GN ILVE OR RC0594.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
Science 293:2093-2098 (2001).
CC -!- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).
CC -!- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-
CC oxopentanoate + L-glutamate.
CC -!- COFACTOR: Pyridoxal phosphate.
CC -!- PATHWAY: Valine and isoleucine biosynthesis.
CC -!- SIMILARITY: Belongs to class-IV of pyridoxal-phosphate-dependent
CC aminotransferases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE005620; AAL03132.1; -
CC PIR; B97774; B97774.
CC InterPro; IPR001544; Aminotrans_IV.
CC Pfam; PF01063; aminotran 4; 1.
CC ProDom; PD001961; Aminotran 4; 1.
CC PROSITE; PS00770; AA TRANSFER CLASS 4; 1.
CC Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
CC Pyridoxal phosphate; Complete proteome.
CC BINDING 155 155 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT

```



SQ SEQUENCE 290 AA; 33243 MW; 86112A254A505177 CRC64;  
Query Match 41.2%; Score 42; DB 1; Length 290;  
Best Local Similarity 55.6%; Pred. No. 27;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
QY 3 KRLLEVGKORLKYAQEE 20  
| : | | : | | | |  
225 KDLGLEVKERLKLQIE 242  
RESULT 22  
PARC ECOLI  
ID PARC ECOLI STANDARD; PRT; 752 AA.  
AC P20082; 069154;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Topoisomerase IV subunit A (EC 5.99.1.1).  
GN PARC OR B3019 OR Z4373 OR ECS3903 OR SF3063 OR S3267.  
OS Escherichia coli.  
OS Escherichia coli O157:H7, and  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562, 83334, 623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=K12;  
RX MEDLINE=9100424; PubMed=2170028;  
RA Kato J.-I., Nishimura Y., Imamura R., Niki H., Hiraga S., Suzuki H.;  
RT "New topoisomerase essential for chromosome segregation in E. coli";  
RL Cell 63:393-404(1990).  
RN [2]  
RP ERRATUM.  
RC SPECIES=E.coli; STRAIN=K12;  
RA Kato J.-I., Nishimura Y., Imamura R., Niki H., Hiraga S., Suzuki H.;  
RL Cell 65:1289-1290(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9279503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocostas K.,  
RA Apodaca J., Anantharanan T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533(2001).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
RN [6]  
RP SEQUENCE OF 724-752 FROM N.A.

RC SPECIES=E.coli;  
RX MEDLINE=92212294; PubMed=1557036;  
RA Coleman J.;  
RT "Characterization of the Escherichia coli gene for  
1-acyl-sn-glycerol-3-phosphate acyltransferase (plac).";  
RL Mol. Gen. Genet. 232:295-303(1992).  
RN [7]  
RP REVISIONS, AND CHARACTERIZATION.  
RC SPECIES=E.coli; STRAIN=K12;  
RX MEDLINE=94043292; PubMed=8227000;  
RA Peng H., Marians K.J.;  
RT "Escherichia coli topoisomerase IV. Purification, characterization,  
subunit structure, and subunit interactions.";  
RL J. Biol. Chem. 268:24481-24481(1993).  
RN [8]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
RX MEDLINE=2272406; PubMed=12384590;  
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding X., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.;  
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
through comparison with genomes of Escherichia coli K12 and O157.";  
RL Nucleic Acids Res. 30:4432-4441(2002).  
RN [9]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
RX MEDLINE=22590274; PubMed=12704152;  
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
RA Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of Shigella  
flexneri serotype 2a strain 2457T.";  
RL Infect. Immun. 71:2775-2786(2003).  
RN [10]  
RP SEQUENCE OF 59-137 FROM N.A.  
RC SPECIES=S.flexneri;  
RX MEDLINE=99056880; PubMed=9867794;  
RA Chu Y.W., Houang E.T.S., Cheng A.F.B.;  
RT "Novel combination of mutations in the DNA gyrase and topoisomerase  
IV genes in laboratory-grown fluoroquinolone-resistant Shigella  
flexneri mutants.";  
RL Antimicrob. Agents Chemother. 42:3051-3052(1998).  
CC -1- FUNCTION: Topoisomerase IV is essential for chromosome  
segregation. It has relaxation of supercoiled DNA activity.  
CC Performs the decatenation events required during the replication  
of a circular DNA molecule.  
CC -1- SUBUNIT: Composed of two subunits: parC and parE.  
CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
CC -1- SIMILARITY: STRONG, WITH THE A SUBUNIT OF GYRASE.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M58408; AAA24297.1; ALT\_INIT.  
CC EMBL; M63491; AAA24396.1;  
CC EMBL; U28377; AAC69187.1;  
CC EMBL; AB000384; AAC76055.1;  
CC EMBL; AF005531; AAG58155.1;  
CC EMBL; AF002563; BAB37326.1;  
CC EMBL; L22025; AAC36840.1;  
CC EMBL; AB015318; AAN44541.1;  
CC EMBL; AB016988; AAPI8353.1;  
CC EMBL; AF065132; AAC17115.1;  
CC PIR; A65089; A65089.

DR PIR; G85961; G85961.  
DR PIR; G91116; G91116.  
DR HSP; P09097; IAB4.  
DR EcoGene; BG10686; parC.  
DR InterPro; IPR005691; DNA\_gyraseA\_C.  
DR InterPro; IPR002205; DNA\_topoisomIV.  
DR InterPro; IPR005742; TopoIV\_A\_Gneg.  
DR Pfam; PF03989; DNA\_gyraseA\_C; 2.  
DR Pfam; PF00521; DNA\_topoisomIV; 1.  
DR ProDom; PD000742; DNA\_topoisomIV; 1.  
DR SMART; SM00434; TOP4C; 1.  
DR TIGRFAMs; TIGR01062; parC\_Gneg; 1.  
DR Topoisomerase; Isomerase; DNA-binding; Complete proteome.  
KW ACT\_SITE 120 120 DNA CLEAVAGE (BY SIMILARITY).  
FT CONFLICT 85 85 A -> P (IN REF. 10).  
SQ SEQUENCE 752 AA; OD4907B96CEE7086 CRC64;  
  
Query Match 41.2%; Score 42; DB 1; Length 752;  
Best Local Similarity 40.0%; Pred. No. 72;  
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
  
QY 1 PYKELRELVGKQRLKYAQEE 20  
DB 694 POSTLTIHVGRKIKLRPEE 713  
  
RESULT 23  
PARC\_SALTY STANDARD; PRT; 752 AA.  
AC P26973;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Topoisomerase IV subunit A (EC 5.99.1.-).  
GN PARC OR STM174  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2;  
RX MEDLINE=92089025; PubMed=1751451;  
RA Luttinger A.L., Springer A.L., Schmid M.B.;  
RA "A cluster of genes that affects nucleoid segregation in Salmonella  
typhimurium";  
RL New Biol. 3:687-697(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grawal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2";  
RL Nature 413:852-856(2001).  
CC -1- FUNCTION: Topoisomerase IV is essential for chromosome  
segregation. It has relaxation of supercoiled DNA activity.  
CC Performs the decatenation events required during the replication  
of a circular DNA molecule.  
CC -1- SUBUNIT: Composed of two subunits: parC and parE.  
CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
CC -1- SIMILARITY: STRONG, WITH THE A SUBUNIT OF GYRASE.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M68936; AAA27180.1; --  
CC EMBL; AB008846; AAL22048.1; --  
CC PIR; A45882; A45882.  
CC HSP; P09097; IAB4.  
CC StyGene; SG10276; parC.  
CC InterPro; IPR006691; DNA\_gyraseA\_C.  
CC InterPro; IPR002205; DNA\_topoisomIV.  
CC InterPro; IPR005742; TopoIV\_A\_Gneg.  
CC Pfam; PF03989; DNA\_gyraseA\_C; 2.  
CC Pfam; PF00521; DNA\_topoisomIV; 1.  
CC ProDom; PD000742; DNA\_topoisomIV; 1.  
CC SMART; SM00434; TOP4C; 1.  
CC TIGRFAMs; TIGR01062; parC\_Gneg; 1.  
CC Topoisomerase; Isomerase; DNA-binding; Complete proteome.  
KW ACT\_SITE 120 120 DNA CLEAVAGE (BY SIMILARITY).  
FT CONFLICT 241 242 MR -> IG (IN REF. 1).  
SQ SEQUENCE 752 AA; 84037 MW; F34FD7FFD20D6760 CRC64;  
  
Query Match 41.2%; Score 42; DB 1; Length 752;  
Best Local Similarity 40.0%; Pred. No. 72;  
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
  
QY 1 PYKELRELVGKQRLKYAQEE 20  
DB 694 POSTLTIHVGRKIKLRPEE 713  
  
RESULT 24  
SYV\_CHLTR STANDARD; PRT; 939 AA.  
AC O84304;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (ValRS).  
GN VALS OR CT302.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D/UW-3/Cx;  
RX MEDLINE=99000809; PubMed=9784136;  
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.;  
RA "Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis";  
RL Science 282:754-759(1998).  
CC -1- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate  
+ L-valyl-tRNA(Val).  
CC -1- SUBUNIT: Monomer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE001302; AAC67895.1; --  
CC PIR; H71532; H71532.  
CC HSP; P96142; IGAX.  
CC InterPro; IPR002300; tRNA-synt\_1a.  
CC InterPro; IPR001412; tRNA-synt\_1.  
CC InterPro; IPR002303; tRNA-synt\_val.  
CC InterPro; IPR009008; ValRS\_1/ERS\_edit.  
CC Pfam; PF00133; tRNA-synt\_1; 1.  
CC

DR PRINTS; PR00986; TRNASYNTHVAL.  
 DR TIGRFAMs; TIGR00422; VALS; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Complete proteome.  
 FT SITE 47 57 "HIGH" REGION.  
 FT SITE 563 567 "KMSKS" REGION.  
 FT BINDING 566 566 ATP (BY SIMILARITY).  
 SQ SEQUENCE 939 AA; 107036 MW; CEB8449DC7BB9066 CRC64;  
 Query Match 41.2%; Score 42; DB 1; Length 939;  
 Best Local Similarity 38.9%; Pred. No. 90;  
 Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 KRLRLVQKQLKYAQEE 20  
 Db 97 RHLKASLGKORTDPSREE 114  
 RESULT 25  
 GBG5 HUMAN  
 ID GBG5 HUMAN STANDARD; PRT; 68 AA.  
 AC P30670; Q61015;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-APR-2004 (Rel. 43, Last annotation update)  
 DE Guanine nucleotide-binding protein G(1)/G(S)/G(O) gamma-5 subunit.  
 GN GNG5 OR GNGT5.  
 OS Homo sapiens (Human);  
 OS Mus musculus (Mouse);  
 OS Rattus norvegicus (Rat); and  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606, 10090, 10116, 9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human;  
 RX MEDLINE=93003227; PubMed=9790912;  
 RA Liu B., Aronson N.N. Jr.;  
 RT "Structure of human G protein Ggamma5 gene GNG5.";  
 RL Biochem. Biophys. Res. Commun. 251:88-94(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human; TISSUE=Blood;  
 RX MEDLINE=9831863; PubMed=9553160;  
 RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,  
 RA He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,  
 RA Wang Y.-X., Chen S.-J., Chen Z.;  
 RT "Identification of genes expressed in human CD34(+) hematopoietic  
 RT stem/progenitor cells by expressed sequence tags and efficient full-  
 RT length cDNA cloning.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human;  
 RX MEDLINE=2238625; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Faisey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Bovine; TISSUE=Liver;  
 RX MEDLINE=92195304; PubMed=1549114;  
 RA Fisher K.J., Aronson N.N. Jr.;  
 RT "Characterization of the cDNA and genomic sequence of a G protein  
 RT gamma subunit (gamma 5).";  
 RL Mol. Cell. Biol. 12:1585-1591(1992).  
 RN [6]  
 RP SEQUENCE.  
 RC SPECIES=Bovine; TISSUE=Spleen;  
 RX MEDLINE=93356792; PubMed=8352779;  
 RA Morishita R., Masuda K., Niwa M., Kato K., Asano T.;  
 RT "Identification of three forms of the gamma subunit of G proteins  
 RT isolated from bovine spleen.";  
 RL Biochem. Biophys. Res. Commun. 194:1221-1227(1993).  
 RN [7]  
 RP SEQUENCE OF 8-53 FROM N.A.  
 RC SPECIES=Mouse; STRAIN=CF-1 / Harlan;  
 RX MEDLINE=97011591; PubMed=8858601;  
 RA Williams C.J., Schultz R.M., Kopf G.S.;  
 RT "G protein gene expression during mouse oocyte growth and maturation,  
 RT and preimplantation embryo development.";  
 RL Mol. Reprod. Dev. 44:315-323(1996).  
 CC -1- FUNCTION: Guanine nucleotide-binding proteins (G proteins) are  
 CC involved as a modulator or transducer in various transmembrane  
 CC signaling systems. The beta and gamma chains are required for the  
 CC GTPase activity, for replacement of GDP by GTP, and for G protein-  
 CC effector interaction.  
 CC -1- SUBUNIT: G proteins are composed of 3 units, alpha, beta and  
 CC gamma.  
 CC -1- TISSUE SPECIFICITY: Expressed in a variety of tissues.  
 CC -1- SIMILARITY: Belongs to the G protein gamma family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF085709; AAC72203.1; -  
 CC EMBL; AF085708; AAC72203.1; JOINED.  
 CC EMBL; AF038955; AAC39869.1; -  
 CC EMBL; AF493873; AAM12587.1; -  
 CC EMBL; BC003563; AAR03563.1; -  
 CC EMBL; M95779; AAA30535.1; -  
 CC EMBL; M95780; AAA41188.1; -  
 CC EMBL; U38498; AAB01729.1; -  
 CC EMBL; B42243; B42243.  
 CC GenBank; U00974; -  
 CC MIM; 600974; -  
 CC MIM; 109164; Gng5.  
 CC InterPro; IPR001770; G-gamma.  
 CC Pfam; PF00631; G-gamma; 1.  
 CC PRINTS; PR00321; GPROTEIN.  
 CC ProDom; PD003783; G-gamma; 1.  
 CC SMART; SM00224; GGL; 1.  
 CC PROSITE; PS00558; G-PROTEIN GAMMA; 1.  
 KW Transducer; Prenylation; Lipoprotein; Multigene family.  
 FT Lipid 65 65 S-geranylgeranyl cysteine

```

FT PROPEP 66 68
SQ SEQUENCE 68 AA; 7318 MW; 9AFA1655863602 CRC64;

Query Match 40.2%; Score 41; DB 1; Length 68;
Best Local Similarity 50.0%; Pred. No. 8.8;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 KEIRLEVGKQRLKYAQ 18
DB 15 QQLRLEAGLRVKVQSQ 30

RESULT 26
GBGA HUMAN
ID -GBGA HUMAN STANDARD; PRT; 68 AA.
AC PS0151;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Guanine nucleotide-binding protein G(i)/G(s)/G(o) gamma-10 subunit.
GN GNG10 OR GNGT10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ISOPRENOID.
RX MEDLINE=95394940; PubMed=7665596;
RA Ray K., Kunsch C., Bonner L.M., Robishaw J.D.;
RT "Isolation of cDNA clones encoding eight different human G protein
gamma subunits, including three novel forms designated the gamma 4,
gamma 10, and gamma 11 subunits."
RL J. Biol. Chem. 270:21765-21771(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Fuhr H.L. III, Ikeda S.R., Aronstam R.S.;
RT "cDNA clones of human proteins involved in signal transduction
sequenced by the Guthrie cDNA resource center (www.cdna.org).";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Laird G.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ustin T.B., Toshiyuki S., Abramson R.D., Mullany S.J.,
Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullany S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Rutgerfeld Y.S.N., Krzywicki M.L., Skalska U., Smalley D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -I- FUNCTION: Guanine nucleotide-binding proteins (G proteins) are
involved as a modulator or transducer in various transmembrane
signaling systems. The beta and gamma chains are required for the
GTPase activity, for replacement of GDP by GTP, and for G protein-
effector interaction. Interacts with beta-1 and beta-2, but not

```

```

CC with beta-3.
CC -I- SUBUNIT: G proteins are composed of 3 units, alpha, beta and
CC gamma.
CC -I- TISSUE SPECIFICITY: Abundantly and ubiquitously expressed.
CC -I- SIMILARITY: Belongs to the G protein gamma family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U31383; AAC50205.1; -
CC EMBL; AF493877; AAM12591.1; -
CC EMBL; AL135787; CAC22155.1; -
CC EMBL; BC010384; AAH10384.1; -
CC EMBL; BC015206; AAH15206.1; -
CC EMBL; BC016319; AAH16319.1; -
CC PIR; I39158; I39158.
CC Genew; HGNC:4402; GNG10.
CC
CC MIM; 604389; -
CC GO; GO:0003927; F:heterotrimeric G-protein GTPase activity; TAS.
CC GO; GO:0007165; P:signal transduction; TAS.
CC InterPro; IPR001770; G-gamma.
CC Pfam; PF00631; G-gamma; 1.
CC PRINTS; PR00321; GPROTEING.
CC ProDom; PD003783; G-gamma; 1.
CC SMART; SM00224; GGL; 1.
CC PROSITE; PS00058; G-PROTEIN GAMMA; 1.
CC Transducer; Prenylation; Lipoprotein; Multigene family.
KW LIPID 65 65 S-geranylgeranyl cysteine.
FT PROPEP 66 68 REMOVED IN MATURE FORM (BY SIMILARITY).
SQ SEQUENCE 68 AA; 7205 MW; E3EC3CA0E0B0D5A CRC64;

Query Match 40.2%; Score 41; DB 1; Length 68;
Best Local Similarity 43.8%; Pred. No. 8.8;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 KEIRLEVGKQRLKYAQ 18
DB 15 EQKLEAGVERIKVQSQ 30

RESULT 27
GST2 HUMAN
ID -GST2 HUMAN STANDARD; PRT; 147 AA.
AC Q99735;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Microsomal glutathione S-transferase 2 (EC 2.5.1.18) (Microsomal GST-
DE 2) (Microsomal GST-II).
DE MGST2 OR GST2.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96355624; PubMed=8703034;
RA Jakobsson P.-J., Mancini J.A., Ford-Hutchinson A.W.;
RT "Identification and characterization of a novel human microsomal
glutathione S-transferase with leukotriene C4 synthase activity and
RT significant sequence identity to 5-lipoxygenase-activating protein
RT and leukotriene C4 synthase."
RL J. Biol. Chem. 271:22203-22210(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

```

RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Matsushita K., Farmer A.A., Rubin G.M., Hong F.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Scherch A., Schein J.E., Jones S.U.M., Marra M.A.;  
 RA "Generation and initial analysis of more than 15,000 full-length  
 RA human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: CAN CATALYZE THE PRODUCTION OF LTC4 FROM LTA4 AND  
 CC -!- REDUCED GLUTATHIONE. CAN CATALYZE THE CONJUGATION OF 1-CHLORO-2-4-  
 CC DINITROBENZENE WITH REDUCED GLUTATHIONE.  
 CC -!- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.  
 CC -!- SUBUNIT: Homotrimer.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- TISSUE SPECIFICITY: Liver, spleen, skeletal muscle, heart,  
 CC adrenals, pancreas, prostate, testis, fetal liver, and fetal  
 CC spleen. Very low expression in lung, brain, placenta and bone  
 CC marrow.  
 CC -!- SIMILARITY: Belongs to the MAPEG family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U77604; AAC51768.1; -;  
 CC EMBL; BC025416; AAH25416.1; -;  
 CC Genew; HGNC:7063; MGST2.  
 CC MIM; 601733; -;  
 CC GO; GO:0005624; C:membrane fraction; TAS.  
 CC GO; GO:0005792; C:microsome; TAS.  
 CC GO; GO:0004364; F:glutathione transferase activity; TAS.  
 CC GO; GO:0006960; P:antimicrobial humoral response (sensu Inver. . .); TAS.  
 CC GO; GO:0007267; P:cell-cell signaling; TAS.  
 CC GO; GO:0006691; P:leukotriene metabolism; TAS.  
 CC GO; GO:0007165; P:signal transduction; TAS.  
 CC InterPro; IPR001446; 5LipoxygenaseP.  
 CC InterPro; IPR001129; MAPEG.  
 CC Pfam; PF01124; MAPEG; 1.  
 CC PRINTS; PR00488; 5LPOXGNASEAP.  
 CC PROSITE; PS01287; FLAP\_GST2\_LTC4S; 1.  
 KW Transferase; Transmembrane; Microsome; Leukotriene biosynthesis.  
 FT TRANSMEM 6 26 POTENTIAL.  
 FT TRANSMEM 59 79 POTENTIAL.  
 FT TRANSMEM 111 131 POTENTIAL.  
 SQ SEQUENCE 147 AA; 16620 MW; DOE89B46885D15EF CRC64;  
  
 QY 7 LEVGKORLKY 16  
 Db 24 LQVGKARLKY 33  
  
 Query Match 40.2%; Score 41; DB 1; Length 147;  
 Best Local Similarity 80.0%; Pred. No. 19;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
 RESULT 28  
 YZ34 AQUAE  
 ID YZ34 AQUAE STANDARD; PRT; 318 AA.

```

066423;
30-MAY-2000 (Rel. 39, Created)
30-MAY-2000 (Rel. 39, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Hypothetical protein AA34.
AA34.
Aquifex aeolicus.
OS
Plasmid ecel.
OC
Bacteria; Aquificae; Aquificaceae; Aquifex.
NCBI_TaxID=63363;
[1]
SEQUENCE FROM N.A.
STRAIN=VF5;
MEDLINE=98196666; PubMed=9537320;
Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., R.,
Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RA
"The complete genome of the hyperthermophilic bacterium Aquifex
RT
aeolicus";
RL
Nature 392:353-358(1998).
CC
-!- SIMILARITY: STRONG, TO A.EOLICUS AA07 AND AA11.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
EMBL; AE000667; AAC07975.1; --
Hypothetical protein; Plasmid; Complete proteome.
SEQUENCE 318 AA; 37360 MW; EDB125C719DE49A3 CRC64;
Query Match 40.2%; Score 41; DB 1; Length 318;
Best Local Similarity 47.4%; Pred. No. 43;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
QY 2 YKELFLEVGKQRLKYAES 20
DB 55 FEEETLAWKQKQKYEES 73
::| | : ||| ||| :||
ATPD MYCTU STANDARD; PRT; 446 AA.
ID ATPD_MYCTU
AC Q10594;
01-OCT-1996 (Rel. 34, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP synthase delta chain [EC 3.6.3.14].
GN ATPH OR RV1307 OR MT1347 OR MTCY373.27 OR MB1339.
OS Mycobacterium tuberculosis, and
OC Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773, 1765;
[1]
SEQUENCE FROM N.A.
SPECIES=My tuberculosis; STRAIN=H37RV;
MEDLINE=98295987; PubMed=96344230;
Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eigmeier K., Gas S., Barry C.B. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA
"Deciphering the biology of Mycobacterium tuberculosis from the
RT
complete genome sequence.";
RL
Nature 393:537-544(1998).
[2]

```

SEQUENCE FROM N.A.  
SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
MEDLINE=22206494; PubMed=12218036;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Emolaeva M., Salzberg S.L.,  
RA Delcher A., Usterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains";  
RL J. Bacteriol. 184:5479-5490(2002).  
RN [3]  
SEQUENCE FROM N.A.  
SPECIES=M.bovis; STRAIN=AF2122/97;  
RX MEDLINE=22709107; PubMed=12788972;  
RA Garnier T., Eighmeier K., Camus J.-C., Medina N., Mansoupe C., Simon S.,  
RA Pryor M., Duchoy S., Gordin S., Lacroix C., Monsemp C., Simon S.,  
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
RT "The complete genome sequence of Mycobacterium bovis";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
CC -!- FUNCTION: This protein seems to be part of the stalk that links  
CF(0) to CF(1). It either transmits conformational changes from  
CF(0) into CF(1) or is implicated in proton conduction.  
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +  
H(+) (Out).  
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic  
core - and CF(0) - the membrane proton channel. CF(1) has five  
subunits: alpha(3), beta(2), gamma(1), delta(1), epsilon(1). CF(0)  
has three main subunits: a, b and c (by similarity).  
CC -!- SIMILARITY: Belongs to the ATPase delta chain family.  
CC -!- SIMILARITY: THIS PROTEIN IS MUCH LONGER THAN THAT OF OTHER  
BACTERIAL DELTA CHAINS, THE C-TERMINAL PART IS HOMOLOGOUS TO DELTA  
CHAINS WHILE THE N-TERMINAL REGION IS SIMILAR TO B/B' SUBUNITS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; Z73419; CAA97740.1; -.  
CC EMBL; AB007008; AAX45609.1; -.  
CC EMBL; BX248338; CAD94200.1; -.  
CC PIR; G70774; G70774.  
CC TIGR; MT1347; -.  
CC TubercuList; Rv1307; -.  
CC InterPro; IPR002146; ATPsynth\_B/B'.sub.  
CC InterPro; IPR000711; ATPsynth\_OSCP.  
CC Pfam; PF00430; ATP-synth B; 1.  
CC Pfam; PF00213; OSCP; 1.  
CC PRINTS; PR00125; ATPASEDELTA.  
CC TIGRFAMs; TIGR01145; ATP\_synth\_delta; 1.  
CC PROSITE; PS00389; ATPASE\_DELTA; FALSE NEG.  
CC K0 Hydroxylase; ATP synthetase; CF(1); Hydrogen ion transport;  
KW Complete proteome.  
FT DOMAIN 1 150 ATP SYNTHASE B/B' SUBUNIT LIKE.  
FT DOMAIN 261 446 ATP SYNTHASE DELTA CHAIN LIKE.  
FT CONFLICT 25 25 L -> V (IN REF. 2).  
SQ SEQUENCE 446 AA; 48805 MW; AF07F26E3B878315 CRC64;  
Query Match 40.2%; Score 41; DB 1; Length 446;  
Best Local Similarity 41.2%; Pred. No. 61;  
Matches 7; Conservative 7; Mismatches 3; Indels 0; Gaps 0;  
QY 3 KRLRLEVGKQRLKYAQE 19  
DB 115 RQLRLELGHESVROARE 131

RESULT 30

MOP HAEIN  
ID -MOP HAEIN STANDARD; PRT; 69 AA.  
AC P45183;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Probable molybdenum-pterin binding protein.  
GN H11370.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kervlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA Klenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd";  
RL Science 269:496-512(1995).  
CC -!- FUNCTION: BINDS ONE MOLE OF MOLYBDENUM PER MOLE OF PROTEIN AND  
CONTAINS A PTERIN (BY SIMILARITY).  
CC -!- SIMILARITY: TO C.PASTEURIANUM MOP PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; U32817; AAC23017.1; -.  
CC PIR; I64119; I64119.  
CC TIGR; H11370; -.  
CC InterPro; IPR004606; Mop.  
CC InterPro; IPR008995; MOP like.  
CC InterPro; IPR005116; TOBE.  
CC Pfam; PF03459; TOBE; 1.  
CC TIGRFAMs; TIGR00638; Mop; 1.  
KW Molybdenum; Complete proteome.  
SQ SEQUENCE 69 AA; 7122 MW; 1D32CE1C4C36310E CRC64;  
Query Match 39.2%; Score 40; DB 1; Length 69;  
Best Local Similarity 80.0%; Pred. No. 13;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 3 KRLRLEVGKQ 12  
DB 45 KRLNLEVGKE 54  
Search completed: May 11, 2004, 13:38:00  
Job time : 8.75 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:28:26 ; Search time 32 Seconds  
(without alignments)  
197.199 Million cell updates/sec

Title: US-09-171-432A-40

Perfect scores: 102

Sequence: 1 PYKELRLEVGKQRLKYAEE 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL 25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mmc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	102	100.0	55	12	Q99972
2	102	100.0	55	12	Q99974
3	102	100.0	55	12	Q99973
4	102	100.0	55	12	Q99978
5	102	100.0	56	12	Q8JYP4
6	102	100.0	56	12	Q67822
7	102	100.0	56	12	Q39872
8	102	100.0	56	12	Q99971
9	102	100.0	56	12	Q99978
10	102	100.0	56	12	Q8JYP2
11	102	100.0	56	12	Q8JYP1
12	102	100.0	56	12	Q99971
13	102	100.0	56	12	Q39867
14	102	100.0	56	12	Q8JYK3
15	102	100.0	56	12	Q918Q3
16	102	100.0	56	12	Q91PB3

17	102	100.0	56	12	Q91PB2
18	102	100.0	56	12	Q99977
19	102	100.0	56	12	Q91P95
20	102	100.0	56	12	Q99HMI
21	102	100.0	56	12	Q39870
22	102	100.0	56	12	Q99HL8
23	102	100.0	56	12	Q91PA1
24	102	100.0	56	12	Q39866
25	102	100.0	56	12	Q8JYK2
26	102	100.0	56	12	Q999U6
27	102	100.0	56	12	Q8JYP7
28	102	100.0	56	12	Q8JYP6
29	102	100.0	56	12	Q91P92
30	102	100.0	56	12	Q9DIX6
31	102	100.0	56	12	Q9DIX9
32	102	100.0	56	12	Q96829
33	102	100.0	56	12	Q91PB8
34	102	100.0	56	12	Q91P97
35	102	100.0	56	12	Q67821
36	102	100.0	56	12	Q9DIX8
37	102	100.0	56	12	Q91P93
38	102	100.0	56	12	Q9DIY4
39	102	100.0	56	12	Q9DIX4
40	102	100.0	56	12	Q8JYT7
41	102	100.0	56	12	Q91PA6
42	102	100.0	56	12	Q99HL7
43	102	100.0	56	12	Q8JYU2
44	102	100.0	56	12	Q91PB1
45	102	100.0	56	12	Q918Q4
46	102	100.0	56	12	Q91PA7
47	102	100.0	56	12	Q91PA8
48	102	100.0	56	12	Q91PD2
49	102	100.0	56	12	Q9DIY0
50	102	100.0	56	12	Q67823
51	102	100.0	56	12	Q91P90
52	102	100.0	56	12	Q39874
53	102	100.0	56	12	Q91PC4
54	102	100.0	56	12	Q91PC7
55	102	100.0	56	12	Q99HMI
56	102	100.0	56	12	Q39869
57	102	100.0	56	12	Q9DIY3
58	102	100.0	56	12	Q9DIX5
59	102	100.0	56	12	Q999V0
60	102	100.0	56	12	Q9DIY1
61	102	100.0	56	12	Q999U5
62	102	100.0	56	12	Q999U9
63	102	100.0	56	12	Q91PC1
64	102	100.0	56	12	Q91PC8
65	102	100.0	56	12	Q918P5
66	102	100.0	56	12	Q91P91
67	102	100.0	56	12	Q39871
68	102	100.0	56	12	Q918P6
69	102	100.0	56	12	Q91P98
70	102	100.0	56	12	Q39868
71	102	100.0	56	12	Q91PC3
72	102	100.0	56	12	Q999T6
73	102	100.0	56	12	Q91PD1
74	102	100.0	56	12	Q918P7
75	102	100.0	56	12	Q8JYP5
76	102	100.0	56	12	Q8JYP3
77	102	100.0	56	12	Q96830
78	102	100.0	56	12	Q91PA3
79	102	100.0	56	12	Q99HM2
80	102	100.0	56	12	Q99HL9
81	102	100.0	56	12	Q999U2
82	102	100.0	56	12	Q999T9
83	102	100.0	56	12	Q91P88
84	102	100.0	56	12	Q999U3
85	102	100.0	56	12	Q91PC2
86	102	100.0	56	12	Q8JYT8
87	102	100.0	56	12	Q8JYP8
88	102	100.0	56	12	Q918Q2
89	102	100.0	56	12	Q918Q1

Q8jvt9 hepatitis a  
Q9ip94 hepatitis a  
Q9diy5 hepatitis a  
Q8jyu0 hepatitis a  
Q9ipb5 hepatitis a  
Q9ipb6 hepatitis a  
Q9ip99 hepatitis a  
Q9dix7 hepatitis a  
Q9ipa5 hepatitis a  
Q99873 hepatitis a  
Q9ipc0 hepatitis a

ALIGNMENTS

RESULT 1  
Q999T2  
ID Q999T2 PRELIMINARY; PRT; 55 AA.  
AC Q999T2;  
DT 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]\_TaxID=12092;  
RP SEQUENCE FROM N.A.  
RC STRAIN=Uru3;  
RA Costa-Mattioli M., Ferre V., Monphoelo S., Garcia L., Colina R.,  
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and  
co-circulation during epidemic outbreaks."  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ306386; CAC29235.1; --  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER\_target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 55  
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;

Query Match 100.0%; Score 102; DB 12; Length 55;  
Best Local Similarity 100.0%; Pred. No. 7.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGVKQRLKYAQEE 20  
DB 35 PYKELRLEVGVKQRLKYAQEE 54

RESULT 2  
Q999T4  
ID Q999T4 PRELIMINARY; PRT; 55 AA.  
AC Q999T4;  
DT 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]\_TaxID=12092;  
RP SEQUENCE FROM N.A.  
RC STRAIN=Chile-3;  
RA Costa-Mattioli M., Ferre V., Monphoelo S., Garcia L., Colina R.,  
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;

"Genetic Variability of Hepatitis C virus reveals heterogeneity and  
co-circulation during epidemic outbreaks."  
Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
EMBL; AJ306386; CAC29233.1; --  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER\_target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 55  
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;

Query Match 100.0%; Score 102; DB 12; Length 55;  
Best Local Similarity 100.0%; Pred. No. 7.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGVKQRLKYAQEE 20  
DB 35 PYKELRLEVGVKQRLKYAQEE 54

RESULT 3  
Q999T3  
ID Q999T3 PRELIMINARY; PRT; 55 AA.  
AC Q999T3;  
DT 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]\_TaxID=12092;  
RP SEQUENCE FROM N.A.  
RC STRAIN=Chile-3;  
RA Costa-Mattioli M., Ferre V., Monphoelo S., Garcia L., Colina R.,  
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and  
co-circulation during epidemic outbreaks."  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ306386; CAC29234.1; --  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER\_target\_S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 55  
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4CB CRC64;

Query Match 100.0%; Score 102; DB 12; Length 55;  
Best Local Similarity 100.0%; Pred. No. 7.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGVKQRLKYAQEE 20  
DB 35 PYKELRLEVGVKQRLKYAQEE 54

RESULT 4  
Q999U8  
ID Q999U8 PRELIMINARY; PRT; 55 AA.  
AC Q999U8;  
DT 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;



```
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Arg-6;
RA Costa-Mattoli M., Ferre V., Monphocho S., Garcia L., Colina R.,
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;
RT "Genetic Variability of Hepatitis A Virus in South America reveals
RT heterogeneity and co-circulation during epidemic outbreaks.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ306370; CAC29219.1; -.
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR InterPro; IPR000886; ER target S.
DR PROSITE; PS00014; ER TARGET; 1.
FT NON_TER 1
FT NON_TER 55
SQ SEQUENCE 55 AA; 6513 MW; 135CB05D46AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 55;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
Db 35 PYKELRLEVGKQRLKYAQEE 54

RESULT 5
Q8JYP4 PRELIMINARY; PRT; 56 AA.
AC Q8JYP4;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Manuvelo;
RA Theamboonlers A., Jantaradsamee P., Poovorawan Y.;
RT "Molecular characterization of Hepatitis A virus infection of an
RT outbreak in the southern part of Thailand.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF507065; AAM33425.1; -.
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR InterPro; IPR000886; ER target S.
DR PROSITE; PS00014; ER TARGET; 1.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
Db 36 PYKELRLEVGKQRLKYAQEE 55

RESULT 6
Q67822 PRELIMINARY; PRT; 56 AA.
ID Q67822
AC Q67822;
```

```
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISR-70;
RA Mendelson E.;
RA Karetnyi Y.V., Shulman L.M., Manor J., Leitner L., Shehab S.;
RT "Genetic classification of hepatitis A virus strains isolated in
RT Israel, based on their VP1/2A nucleotide sequence.";
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z77247; CAB01040.1; -.
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0431; PQ0431.
DR InterPro; IPR000886; ER target S.
DR PROSITE; PS00014; ER TARGET; 1.
FT NON_TER 1
FT CHAIN 1 >28 CAPSID PROTEIN VP1.
FT CHAIN 29 >56 NONSTRUCTURAL PROTEIN NS2A.
FT NON_TER 56
FT NON_TER 56
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
Db 36 PYKELRLEVGKQRLKYAQEE 55

RESULT 7
O39872 PRELIMINARY; PRT; 56 AA.
ID O39872;
AC O39872;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VDM;
RX MEDLINE=97247817; PubMed=9093940;
RA Taylor M.B.;
RT "Molecular epidemiology of South African strains of hepatitis A virus:
RT 1982-1996.";
RL J. Med. Virol. 51:273-279(1997).
DR EMBL; U68697; AAB53593.1; -.
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR InterPro; IPR000886; ER target S.
DR PROSITE; PS00014; ER TARGET; 1.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
```

```
Db 36 PYKELRLEVKGKRLKYAQEE 55
|||||
ID Q999T1 PRELIMINARY; PRT; 56 AA.
AC Q999T1, 2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE Polypeptide (Fragment)
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H202;
RA Costa-Mattioli M., Ferre V., Monpholo S., Garcia L., Colina R.,
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;
RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and
RT co-circulation during epidemic outbreaks.";
RL Submitted (FE2-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ306387; CAC29236.1; -
DR PIR; PQ0427; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6602 MW; CE935CE05D46AEF4 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVKGKRLKYAQEE 20
|||||
Db 35 PYKELRLEVKGKRLKYAQEE 54
|||||
ID Q999T8 PRELIMINARY; PRT; 56 AA.
AC Q999T8, 2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE Polypeptide (Fragment)
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Chile-4;
RA Costa-Mattioli M., Ferre V., Monpholo S., Garcia L., Colina R.,
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;
RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and
RT co-circulation during epidemic outbreaks.";
RL Submitted (FE2-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ306380; CAC29229.1; -
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVKGKRLKYAQEE 20
|||||
Db 35 PYKELRLEVKGKRLKYAQEE 54
|||||
ID Q999T8 PRELIMINARY; PRT; 56 AA.
AC Q999T8, 2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE Polypeptide (Fragment)
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Chile-4;
RA Costa-Mattioli M., Ferre V., Monpholo S., Garcia L., Colina R.,
RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;
RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and
RT co-circulation during epidemic outbreaks.";
RL Submitted (FE2-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ306380; CAC29229.1; -
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;
```

```
Query Match 100.0%; Score 102; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVKGKRLKYAQEE 20
|||||
Db 36 PYKELRLEVKGKRLKYAQEE 55
|||||
ID Q8JYP2 PRELIMINARY; PRT; 56 AA.
AC Q8JYP2, 2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DE Polypeptide (Fragment)
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HAVAO;
RA Theamboonlers A., Jantaradamee P., Poovorawan Y.;
RT "Molecular characterization of Hepatitis A virus infection of an
RT outbreak in the southern part of Thailand.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF507067; AAM33427.1; -
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR InterPro; IPR000886; ER target S.
DR PROSITE; PS00014; ER_TARGET; 1.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVKGKRLKYAQEE 20
|||||
Db 36 PYKELRLEVKGKRLKYAQEE 55
|||||
ID Q8JYP1 PRELIMINARY; PRT; 56 AA.
AC Q8JYP1, 2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DE Polypeptide (Fragment)
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Maya;
RA Theamboonlers A., Jantaradamee P., Poovorawan Y.;
RT "Molecular characterization of Hepatitis A virus infection of an
RT outbreak in the southern part of Thailand.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF507068; AAM33428.1; -
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR InterPro; IPR000886; ER target S.
```

DR PROSITE; PS00014; ER\_TARGET; 1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;  
 Query Match 100.0%; Score 102; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PYKELRLEVGVKQRLKYAQEE 20  
 DB 36 PYKELRLEVGVKQRLKYAQEE 55  
 RESULT 12  
 Q999U1 PRELIMINARY; PRT; 56 AA.  
 AC Q999U1;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polypeptide (Fragment)  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Arg-23;  
 RA Costa-Mattioli M., Ferre V., Monphoelo S., Garcia L., Colina R.,  
 RA Billaudel S., Vega I., Perez-Bercoff R., Cristina J.;  
 RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and  
 RT co-circulation during epidemic outbreaks."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ306377; CAC29226.1;  
 DR PIR; PQ0427; PQ0427.  
 DR PIR; PQ0428; PQ0428.  
 DR PIR; PQ0430; PQ0430.  
 DR PIR; PQ0431; PQ0431.  
 DR InterPro; IPR000896; ER target S.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;  
 Query Match 100.0%; Score 102; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PYKELRLEVGVKQRLKYAQEE 20  
 DB 36 PYKELRLEVGVKQRLKYAQEE 55  
 RESULT 13  
 Q39867 PRELIMINARY; PRT; 56 AA.  
 AC Q39867;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polypeptide (Fragment)  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JVR;  
 RX MEDLINE=97247817; PubMed=9093940;  
 RA Taylor M.B.;  
 RT "Molecular epidemiology of South African strains of hepatitis A virus:  
 RT 1982-1996."

J. Med. Virol. 51:273-279(1997).  
 RL EMBL; U68692; AAB53588.1;  
 DR PIR; PQ0427; PQ0427.  
 DR PIR; PQ0428; PQ0428.  
 DR PIR; PQ0431; PQ0431.  
 DR InterPro; IPR000886; ER target S.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;  
 Query Match 100.0%; Score 102; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PYKELRLEVGVKQRLKYAQEE 20  
 DB 36 PYKELRLEVGVKQRLKYAQEE 55  
 RESULT 14  
 Q8JYK3 PRELIMINARY; PRT; 56 AA.  
 AC Q8JYK3;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polypeptide (Fragment)  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Yeengor 15;  
 RA Theamboonlers A., Jantaradsamee P., Poovorawan Y.;  
 RT "Molecular characterization of Hepatitis A virus infection of an  
 RT outbreak in the southern part of Thailand."  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF509835; AAM34763.1;  
 DR PIR; PQ0427; PQ0427.  
 DR PIR; PQ0428; PQ0428.  
 DR PIR; PQ0430; PQ0430.  
 DR PIR; PQ0431; PQ0431.  
 DR InterPro; IPR000886; ER target S.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 FT NON\_TER 1  
 FT NON\_TER 56  
 SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;  
 Query Match 100.0%; Score 102; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PYKELRLEVGVKQRLKYAQEE 20  
 DB 36 PYKELRLEVGVKQRLKYAQEE 55  
 RESULT 15  
 Q918Q3 PRELIMINARY; PRT; 56 AA.  
 AC Q918Q3;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polypeptide (Fragment)  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=RJ-005;  
RA de Paula V.S., Baptista M.L., Lampe E., Niel C., Gaspar A.M.C.;  
RT "Characterization of hepatitis A virus isolates from subgenotypes IA  
and IB in Rio de Janeiro, Brazil.";  
RL J. Med. Virol. 0:0-0(2001).  
DR EMBL: AF410391; AAL10177.1; --  
DR PIR: PQ0427; PQ0427.  
DR PIR: PQ0428; PQ0428.  
DR PIR: PQ0431; PQ0431.  
DR InterPro: IPR000886; ER target S.  
DR PROSITE: PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438CS1846AEF4A4 CRC64;  
  
Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
DB 36 PYKELRLEVGKQRLKYAQEE 55  
  
RESULT 16  
Q9IPB3 PRELIMINARY; PRT; 56 AA.  
AC Q9IPB3;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment)  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RC STRAIN=A306;  
RT "hepatitis A virus VP1/2A junction.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB046897; BAB08054.1; --  
DR PIR: PQ0427; PQ0427.  
DR PIR: PQ0428; PQ0428.  
DR PIR: PQ0430; PQ0430.  
DR PIR: PQ0431; PQ0431.  
DR InterPro: IPR000886; ER target S.  
DR PROSITE: PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;  
  
Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
DB 36 PYKELRLEVGKQRLKYAQEE 55  
  
RESULT 17  
Q9IPB2 PRELIMINARY; PRT; 56 AA.  
AC Q9IPB2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment)  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.

OX NCBI\_TaxID=12092;  
RN [1]  
RC STRAIN=A307;  
RT "hepatitis A virus VP1/2A junction.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB046898; BAB08055.1; --  
DR PIR: PQ0427; PQ0427.  
DR PIR: PQ0428; PQ0428.  
DR PIR: PQ0430; PQ0430.  
DR PIR: PQ0431; PQ0431.  
DR InterPro: IPR000886; ER target S.  
DR PROSITE: PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;  
  
Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
DB 36 PYKELRLEVGKQRLKYAQEE 55  
  
RESULT 18  
Q999U7 PRELIMINARY; PRT; 56 AA.  
AC Q999U7;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment)  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RC STRAIN=Uru7;  
RT "hepatitis A virus VP1/2A junction.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ306371; CAC29220.1; --  
DR PIR: PQ0427; PQ0427.  
DR PIR: PQ0428; PQ0428.  
DR PIR: PQ0430; PQ0430.  
DR PIR: PQ0431; PQ0431.  
DR InterPro: IPR000886; ER target S.  
DR PROSITE: PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;  
  
Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
DB 36 PYKELRLEVGKQRLKYAQEE 55  
  
RESULT 19  
Q9IP95 PRELIMINARY; PRT; 56 AA.  
AC Q9IP95;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1] \_  
RP SEQUENCE FROM N.A.  
RC STRAIN=A7;  
RA Fujiwara K.;  
RT "hepatitis A virus VP1/2A junction."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB046915; BAB08072.1; -  
DR PIR; PQ0427; PQ0428.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 1  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;  
  
Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 36 PYKELRLEVGKQRLKYAQEE 55  
  
RESULT 20  
Q99HM1 PRELIMINARY; PRT; 56 AA.  
AC Q99HM1;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Capsid protein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1] \_  
RP SEQUENCE FROM N.A.  
RC STRAIN=KU94-420;  
RA Song J.W., Byun K.S., Kim J.H., Song K.J., Baek L.J., Park S.H.,  
RA Kwon O.S.;  
RT "Molecular epidemiology of hepatitis A virus in Korea."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF234877; AAK00782.1; -  
DR PIR; PQ0427; PQ0428.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 1  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;  
  
Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 36 PYKELRLEVGKQRLKYAQEE 55  
  
RESULT 21  
Q99870 PRELIMINARY; PRT; 56 AA.  
ID Q99870

039870;  
AC 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1] \_  
RP SEQUENCE FROM N.A.  
RC STRAIN=406808;  
RX MEDLINE=97247817; PubMed=9093940;  
RA Taylor M.B.;  
RT "Molecular epidemiology of South African strains of hepatitis A virus."  
RL J. Med. Virol. 51:273-279 (1997).  
DR EMBL; U68695; AAB53591.1; -  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 1  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;  
  
Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 36 PYKELRLEVGKQRLKYAQEE 55  
  
RESULT 22  
Q99HL8 PRELIMINARY; PRT; 56 AA.  
AC Q99HL8;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Capsid protein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1] \_  
RP SEQUENCE FROM N.A.  
RC STRAIN=KU98-54;  
RA Song J.W., Byun K.S., Kim J.H., Song K.J., Baek L.J., Park S.H.,  
RA Kwon O.S.;  
RT "Molecular epidemiology of hepatitis A virus in Korea."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF234880; AAK00785.1; -  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
FT NON\_TER 1 1  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;  
  
Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 36 PYKELRLEVGKQRLKYAQEE 55  
  
RESULT 23  
Q99870 PRELIMINARY; PRT; 56 AA.  
ID Q99870

Db 36 PYKELRLEVKGKRLKYAQEE 55

RESULT 23

Q91PA1 Q91PA1 PRELIMINARY; PRT; 56 AA.

AC Q91PA1

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Polypeptide (Fragment)

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OC NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AS8;

RA Fujiwara K.;

RT "Hepatitis A virus VP1/2A junction.";

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB046909; BAB08066.1; -

DR PIR; PQ0427; PQ0427.

DR PIR; PQ0428; PQ0428.

DR PIR; PQ0430; PQ0430.

DR PIR; PQ0431; PQ0431.

DR InterPro; IPR000886; ER\_target\_S.

DR PROSITE; PS00014; ER\_TARGET; 1.

FT NON\_TER 1

FT NON\_TER 56

SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;

Best Local Similarity 100.0%; Pred. No. 7.7e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVKGKRLKYAQEE 20

Db 36 PYKELRLEVKGKRLKYAQEE 55

RESULT 24

O39866 O39866 PRELIMINARY; PRT; 56 AA.

AC O39866

DT 01-JAN-1998 (TRENBLrel. 05, Created)

DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Polypeptide (Fragment)

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OC NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=96001190;

RX MEDLINE=97247817; PubMed=9093940;

RA Taylor M.B.;

RT "Molecular epidemiology of South African strains of hepatitis A virus: 1982-1996.";

RL J. Med. Virol. 51:273-279(1997).

DR EMBL; U68690; AAB53586.1; -

DR PIR; PQ0427; PQ0427.

DR PIR; PQ0428; PQ0428.

DR PIR; PQ0431; PQ0431.

DR InterPro; IPR000886; ER\_target\_S.

DR PROSITE; PS00014; ER\_TARGET; 1.

FT NON\_TER 1

FT NON\_TER 56

SQ SEQUENCE 56 AA; 6614 MW; 8438C51846AEF4A4 CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;

Best Local Similarity 100.0%; Pred. No. 7.7e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVKGKRLKYAQEE 20

Db 36 PYKELRLEVKGKRLKYAQEE 55

RESULT 25

O8JYK2 O8JYK2 PRELIMINARY; PRT; 56 AA.

AC O8JYK2

DT 01-OCT-2002 (TRENBLrel. 22, Created)

DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Polypeptide (Fragment)

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OC NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Yeengor 16;

RA Theamboonlers A.; Jantaradsamee P.; Poovorawan Y.;

RT "Molecular characterization of Hepatitis A virus infection of an outbreak in the southern part of Thailand.";

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF509836; AAM34764.1; -

DR PIR; PQ0427; PQ0427.

DR PIR; PQ0428; PQ0428.

DR PIR; PQ0430; PQ0430.

DR PIR; PQ0431; PQ0431.

DR InterPro; IPR000886; ER\_target\_S.

DR PROSITE; PS00014; ER\_TARGET; 1.

FT NON\_TER 1

FT NON\_TER 56

SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;

Best Local Similarity 100.0%; Pred. No. 7.7e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVKGKRLKYAQEE 20

Db 36 PYKELRLEVKGKRLKYAQEE 55

RESULT 26

O999U6 O999U6 PRELIMINARY; PRT; 56 AA.

AC O999U6

DT 01-JUN-2001 (TRENBLrel. 17, Created)

DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Polypeptide (Fragment)

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OC NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ARG-4;

RA Costa-Wattoli M.; Perre V.; Mompheolo S.; Garcia L.; Colina R.;

RA Billaudel S.; Vega I.; Perez-Bercoff R.; Cristina J.;

RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and co-circulation during epidemic outbreaks.";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ306372; CAC29221.1; -

DR PIR; PQ0427; PQ0427.

DR PIR; PQ0428; PQ0428.

DR PIR; PQ0430; PQ0430.

DR PIR; PQ0431; PQ0431.

DR InterPro; IPR000886; ER\_target\_S.

DR PROSITE; PS00014; ER\_TARGET; 1.

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVKGKRLKYAQEE 20

Db 36 PYKELRLEVKGKRLKYAQEE 55

RESULT 25

O8JYK2 O8JYK2 PRELIMINARY; PRT; 56 AA.

AC O8JYK2

DT 01-OCT-2002 (TRENBLrel. 22, Created)

DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Polypeptide (Fragment)

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OC NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Yeengor 16;

RA Theamboonlers A.; Jantaradsamee P.; Poovorawan Y.;

RT "Molecular characterization of Hepatitis A virus infection of an outbreak in the southern part of Thailand.";

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF509836; AAM34764.1; -

DR PIR; PQ0427; PQ0427.

DR PIR; PQ0428; PQ0428.

DR PIR; PQ0430; PQ0430.

DR PIR; PQ0431; PQ0431.

DR InterPro; IPR000886; ER\_target\_S.

DR PROSITE; PS00014; ER\_TARGET; 1.

FT NON\_TER 1

FT NON\_TER 56

SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;

Best Local Similarity 100.0%; Pred. No. 7.7e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVKGKRLKYAQEE 20

Db 36 PYKELRLEVKGKRLKYAQEE 55

RESULT 26

O999U6 O999U6 PRELIMINARY; PRT; 56 AA.

AC O999U6

DT 01-JUN-2001 (TRENBLrel. 17, Created)

DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Polypeptide (Fragment)

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OC NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ARG-4;

RA Costa-Wattoli M.; Perre V.; Mompheolo S.; Garcia L.; Colina R.;

RA Billaudel S.; Vega I.; Perez-Bercoff R.; Cristina J.;

RT "Genetic Variability of Hepatitis C virus reveals heterogeneity and co-circulation during epidemic outbreaks.";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ306372; CAC29221.1; -

DR PIR; PQ0427; PQ0427.

DR PIR; PQ0428; PQ0428.

DR PIR; PQ0430; PQ0430.

DR PIR; PQ0431; PQ0431.

DR InterPro; IPR000886; ER\_target\_S.

DR PROSITE; PS00014; ER\_TARGET; 1.

```
FT NON_TER 1 1
SQ NON_TER 56
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
DB 36 PYKELRLEVGKQRLKYAQEE 55

RESULT 27
Q8JYP7 PRELIMINARY; PRT; 56 AA.
AC Q8JYP7;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Polypeptide (Fragment)
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=stMA;
RA Theamboonlers A., Jantaradsamee P., Poovorawan Y.;
RT "Molecular characterization of Hepatitis A virus infection of an
outbreak in the southern part of Thailand."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF507062; AM33422.1; -
DR PIR; PQ0427; PQ0428.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR InterPro; IPR000886; ER_target_S.
DR PROSITE; PS00014; ER_TARGET; 1.
FT NON_TER 1 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
DB 36 PYKELRLEVGKQRLKYAQEE 55

RESULT 28
Q8JYP6 PRELIMINARY; PRT; 56 AA.
AC Q8JYP6;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Polypeptide (Fragment)
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB;
RA Theamboonlers A., Jantaradsamee P., Poovorawan Y.;
RT "Molecular characterization of Hepatitis A virus infection of an
outbreak in the southern part of Thailand."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF507063; AM33423.1; -
DR PIR; PQ0427; PQ0427.
```

```
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR InterPro; IPR000886; ER_target_S.
DR PROSITE; PS00014; ER_TARGET; 1.
FT NON_TER 1 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
DB 36 PYKELRLEVGKQRLKYAQEE 55

RESULT 29
Q91P92 PRELIMINARY; PRT; 56 AA.
AC Q91P92;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Polypeptide (Fragment)
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A75;
RA Fujiwara K.;
RT "Hepatitis A virus VP1/2A junction."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB046918; BAB08075.1; -
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR InterPro; IPR000886; ER_target_S.
DR PROSITE; PS00014; ER_TARGET; 1.
FT NON_TER 1 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6642 MW; 465CF51846AEF4BC CRC64;

Query Match 100.0%; Score 102; DB 12; Length 56;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20
DB 36 PYKELRLEVGKQRLKYAQEE 55

RESULT 30
Q9DIX6 PRELIMINARY; PRT; 56 AA.
AC Q9DIX6;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE VP1 capsid protein and P2A protease (Fragment)
GN VP1 AND P2A.
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OX NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CU-263P;
RX MEDLINE=21259955; PubMed=11360240;
```

RA Diaz B.I., Sariol C.A., Normann A., Rodriguez L.A., Flehmig B.;  
RT "Genetic relatedness of Cuban HAV wild-type isolates.";  
RL J. Med. Virol. 64:96-103 (2001).  
DR EMBL; AJ245531; CAC17884.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR InterPro; IPR000886; ER target S.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
KW Protease.  
FT NON\_TER 1 1  
FT NON\_TER 56 56  
SQ SEQUENCE 56 AA; 6628 MW; 465CF4B35C1EF4BC CRC64;  
  
Query Match 100.0%; Score 102; DB 12; Length 56;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 PYKELRLEVGVKQLKYAQEE 20  
|||  
Db 36 PYKELRLEVGVKQLKYAQEE 55  
|||  
  
Search completed: May 11, 2004, 13:40:23  
Job time : 34 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:37:37 ; Search time 34.25 Seconds  
(without alignments)  
162.083 Million cell updates/sec

Title: US-09-171-432A-40

Perfect score: 102  
Sequence: 1 PYKELRLEVGKQLKQAE 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 27756755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

#### Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US05\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	102	100.0	20	10	US-09-171-432A-40
2	102	100.0	352	14	US-10-272-459-45
3	102	100.0	836	14	US-10-272-459-40
4	102	100.0	980	14	US-10-272-459-41
5	102	100.0	2227	9	US-09-929-955-12
6	102	100.0	2227	13	US-10-104-966-12
7	102	100.0	2227	13	US-10-135-988-2
8	102	100.0	2227	13	US-10-135-988-4
9	102	100.0	2227	13	US-10-135-988-6
10	67	65.7	20	10	US-09-171-432A-39
11	46.5	45.6	65	11	US-09-864-408A-140
12	45	44.1	20	10	US-09-171-432A-41
13	45	44.1	25	10	US-09-171-432A-47
14	45	44.1	214	15	US-10-289-762-108
15	44	43.1	211	14	US-10-238-075-1284

16	44	43.1	270	12	US-10-424-599-252284	Sequence 252284,
17	44	43.1	720	12	US-10-312-373-1	Sequence 1, Appli
18	44	43.1	720	15	US-10-390-585-16	Sequence 16, Appli
19	44	43.1	720	16	US-10-055-569A-35	Sequence 35, Appli
20	44	43.1	720	16	US-10-055-569A-36	Sequence 36, Appli
21	44	43.1	720	16	US-10-055-569A-37	Sequence 37, Appli
22	43.5	42.6	1273	12	US-10-363-616-325	Sequence 325, App
23	43	42.2	120	14	US-10-238-075-939	Sequence 939, App
24	43	42.2	136	12	US-10-424-599-285976	Sequence 285976,
25	43	42.2	178	14	US-10-091-007-134	Sequence 124, App
26	42	41.2	90	12	US-10-424-599-216526	Sequence 216526,
27	42	41.2	156	12	US-10-424-599-226551	Sequence 226551,
28	42	41.2	207	14	US-10-238-075-1123	Sequence 1123, Ap
29	42	41.2	443	12	US-10-424-599-212457	Sequence 212457,
30	42	41.2	736	12	US-10-282-122A-76680	Sequence 76680, A
31	42	41.2	752	12	US-10-282-122A-56689	Sequence 56689, A
32	42	41.2	1054	12	US-10-001-885-108	Sequence 108, App
33	42	41.2	1104	12	US-10-425-114-57274	Sequence 57274, A
34	42	41.2	1122	12	US-10-412-699B-1582	Sequence 1582, Ap
35	42	41.2	4131	15	US-10-369-493-5136	Sequence 5136, Ap
36	41.5	40.7	352	15	US-10-369-493-3258	Sequence 3258, Ap
37	41.5	40.7	733	12	US-10-424-599-243766	Sequence 243766,
38	41	40.2	34	12	US-10-424-599-210256	Sequence 210256,
39	41	40.2	46	9	US-09-864-761-34339	Sequence 34339, A
40	41	40.2	68	14	US-10-161-941-18	Sequence 18, Appli
41	41	40.2	68	14	US-10-161-941-19	Sequence 19, Appli
42	41	40.2	69	9	US-09-982-809-5	Sequence 5, Appli
43	41	40.2	69	9	US-09-982-809-7	Sequence 7, Appli
44	41	40.2	77	12	US-10-424-599-189611	Sequence 189611,
45	41	40.2	79	9	US-09-925-301-877	Sequence 877, App
46	41	40.2	147	12	US-10-294-934-488	Sequence 488, App
47	41	40.2	147	14	US-10-247-671-133	Sequence 133, App
48	41	40.2	169	12	US-10-424-599-270895	Sequence 270895,
49	41	40.2	187	9	US-09-925-299-881	Sequence 881, App
50	41	40.2	187	10	US-09-925-299-881	Sequence 881, App
51	41	40.2	187	15	US-10-264-043-2421	Sequence 2421, Ap
52	41	40.2	337	12	US-10-424-599-225698	Sequence 225698,
53	41	40.2	397	14	US-10-168-425-13	Sequence 13, Appli
54	41	40.2	446	12	US-10-282-122A-62827	Sequence 62827, A
55	41	40.2	446	12	US-10-282-122A-64532	Sequence 64532, A
56	41	40.2	666	16	US-10-389-566-589	Sequence 589, App
57	41	40.2	767	15	US-10-369-493-9794	Sequence 9794, Ap
58	41	40.2	859	12	US-10-282-122A-65982	Sequence 65982, A
59	41	40.2	1137	12	US-10-425-114-63876	Sequence 63876, A
60	40.5	39.7	215	12	US-10-424-599-149292	Sequence 149292,
61	40	39.2	24	9	US-09-789-404-15	Sequence 15, Appli
62	40	39.2	171	12	US-10-424-599-160390	Sequence 160390,
63	40	39.2	182	14	US-10-191-513A-20	Sequence 20, Appli
64	40	39.2	189	9	US-09-847-209-2	Sequence 2, Appli
65	40	39.2	189	9	US-09-847-209-4	Sequence 4, Appli
66	40	39.2	189	15	US-10-437-581-2	Sequence 2, Appli
67	40	39.2	219	14	US-10-191-513A-19	Sequence 19, Appli
68	40	39.2	219	9	US-09-925-297-661	Sequence 661, App
69	40	39.2	272	9	US-10-191-513A-14	Sequence 14, Appli
70	40	39.2	287	14	US-10-191-513A-15	Sequence 15, Appli
71	40	39.2	288	14	US-10-425-114-45339	Sequence 45339, A
72	40	39.2	289	12	US-10-191-513A-39	Sequence 39, Appli
73	40	39.2	347	14	US-10-191-513A-40	Sequence 40, Appli
74	40	39.2	360	14	US-10-191-513A-39	Sequence 39, Appli
75	40	39.2	392	14	US-10-181-071-5	Sequence 5, Appli
76	40	39.2	419	9	US-09-948-774-2	Sequence 2, Appli
77	40	39.2	419	9	US-09-738-626-4556	Sequence 4556, Ap
78	40	39.2	425	12	US-10-296-115-1340	Sequence 1240, Ap
79	40	39.2	444	12	US-10-231-956A-69	Sequence 69, Appli
80	40	39.2	444	14	US-10-262-617-3	Sequence 3, Appli
81	40	39.2	444	14	US-10-191-513A-12	Sequence 12, Appli
82	40	39.2	444	14	US-10-191-513A-42	Sequence 42, Appli
83	40	39.2	444	16	US-10-429-160-22	Sequence 22, Appli
84	40	39.2	470	15	US-10-369-493-23388	Sequence 23388, A
85	40	39.2	498	15	US-10-289-762-143	Sequence 143, App
86	40	39.2	501	12	US-10-335-977-8259	Sequence 8259, Ap
87	40	39.2	505	12	US-10-335-977-8360	Sequence 8260, Ap
88	40	39.2	578	12	US-10-282-122A-72476	Sequence 72476, A

89 40 39.2 685 10 US-09-877-843-58 Sequence 58, Appl  
90 40 39.2 685 10 US-09-877-843-59 Sequence 59, Appl  
91 40 39.2 764 9 US-09-738-973-67 Sequence 67, Appl  
92 40 39.2 764 9 US-09-854-133-67 Sequence 67, Appl  
93 40 39.2 764 14 US-10-144-649A-67 Sequence 67, Appl  
94 40 39.2 775 14 US-10-114-153-58 Sequence 58, Appl  
95 40 39.2 794 9 US-09-789-404-2 Sequence 2, Appl  
96 40 39.2 795 12 US-10-282-122A-64231 Sequence 64231, A  
97 40 39.2 797 15 US-10-424-233-16 Sequence 16, Appl  
98 40 39.2 816 15 US-10-369-493-3409 Sequence 3409, A  
99 40 39.2 939 12 US-10-282-122A-55211 Sequence 55211, A  
100 40 39.2 1028 14 US-10-153-668-192 Sequence 192, App

# ALIGNMENTS

## RESULT 1

US-09-171-432A-40  
; Sequence 40, Application US/09171432A  
; Publication No. US200301841  
; GENERAL INFORMATION:  
; APPLICANT: Fields, Howard A.  
; APPLICANT: Khudyakov, Yuri E.  
; TITLE OF INVENTION: Antigenically Reactive Regions of the  
; TITLE OF INVENTION: Hepatitis A Virus Polyprotein  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Kilpatrick Stockton LLP  
; STREET: 3424 Peachtree Road, N.E.  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30326

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/171,432A  
; FILING DATE: 23-NOV-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Greene, Jamie L.  
; REGISTRATION NUMBER: 32,467  
; REFERENCE/DOCKET NUMBER: 03063-0231US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 949-2400  
; TELEFAX: (404) 949-2499  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..20  
; OTHER INFORMATION: /label= YK-1316

## US-09-171-432A-40

Query Match 100.0%; Score 102; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
|||  
Db 1 PYKELRLEVGKQRLKYAQEE 20  
|||

## US-09-171-432A-40

Query Match 100.0%; Score 102; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PYKELRLEVGKQRLKYAQEE 20  
|||  
Db 1 PYKELRLEVGKQRLKYAQEE 20  
|||

## RESULT 2

US-10-272-459-45  
; Sequence 45, Application US/10272459  
; Publication No. US20030124517A1  
; GENERAL INFORMATION:  
; APPLICANT: PICHUANES, Sergio  
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; FILE REFERENCE: PP17955.002 / 2301-17955  
; CURRENT APPLICATION NUMBER: US/10/272,459  
; CURRENT FILING DATE: 2002-10-15  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 352  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein  
; OTHER INFORMATION: of 38.8 kDa  
US-10-272-459-45

Query Match 100.0%; Score 102; DB 14; Length 352;  
Best Local Similarity 100.0%; Pred. No. 3.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
|||  
Db 171 PYKELRLEVGKQRLKYAQEE 190  
|||

## RESULT 3

US-10-272-459-40  
; Sequence 40, Application US/10272459  
; Publication No. US20030124517A1  
; GENERAL INFORMATION:  
; APPLICANT: PICHUANES, Sergio  
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; FILE REFERENCE: PP17955.002 / 2301-17955  
; CURRENT APPLICATION NUMBER: US/10/272,459  
; CURRENT FILING DATE: 2002-10-15  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 40  
; LENGTH: 836  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein  
; OTHER INFORMATION: of 94 kDa  
US-10-272-459-40

Query Match 100.0%; Score 102; DB 14; Length 836;  
Best Local Similarity 100.0%; Pred. No. 8.3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
|||  
Db 799 PYKELRLEVGKQRLKYAQEE 818  
|||

## RESULT 4

US-10-272-459-41  
; Sequence 41, Application US/10272459  
; Publication No. US20030124517A1  
; GENERAL INFORMATION:  
; APPLICANT: PICHUANES, Sergio  
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; FILE REFERENCE: PP17955.002 / 2301-17955

; CURRENT APPLICATION NUMBER: US/10/272,459  
; CURRENT FILING DATE: 2002-10-15  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 41  
; LENGTH: 980  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein  
; OTHER INFORMATION: of 115.5 kDa  
US-10-272-459-41

Query Match 100.0%; Score 102; DB 14; Length 980;  
Best Local Similarity 100.0%; Pred. No. 9.9e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 799 PYKELRLEVGKQRLKYAQEE 818

#### RESULT 5

US-09-929-955-12  
; Sequence 12, Application US/09929955  
; Patent No. US20020136740A1  
; GENERAL INFORMATION:  
; APPLICANT: Matti Sallberg  
; APPLICANT: Catharina Hultgren  
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: TRIPEP.23AUS2  
; CURRENT APPLICATION NUMBER: US/09/929,955  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/705,547  
; PRIOR FILING DATE: 2000-11-03  
; PRIOR APPLICATION NUMBER: 60/229,175  
; PRIOR FILING DATE: 2000-08-29  
; PRIOR APPLICATION NUMBER: 60/225,767  
; PRIOR FILING DATE: 2000-08-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Hepatitis A virus sequence  
US-09-929-955-12

Query Match 100.0%; Score 102; DB 9; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 799 PYKELRLEVGKQRLKYAQEE 818

#### RESULT 6

US-10-104-966-12  
; Sequence 12, Application US/10104966  
; Publication No. US2002055124A1  
; GENERAL INFORMATION:  
; APPLICANT: Matti Sallberg  
; APPLICANT: Catharina Hultgren  
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: TRIPEP.23AUSC1  
; CURRENT APPLICATION NUMBER: US/10/104,966  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/705,547  
; PRIOR FILING DATE: 2000-11-03

; PRIOR APPLICATION NUMBER: 60/229,175  
; PRIOR FILING DATE: 2000-08-29  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Hepatitis A virus sequence  
US-10-104-966-12

Query Match 100.0%; Score 102; DB 13; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 799 PYKELRLEVGKQRLKYAQEE 818

#### RESULT 7

US-10-135-988-2  
; Sequence 2, Application US/10135988  
; Publication No. US20020176869A1  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-10-135-988-2

Query Match 100.0%; Score 102; DB 13; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
Db 799 PYKELRLEVGKQRLKYAQEE 818

#### RESULT 8

US-10-135-988-4  
; Sequence 4, Application US/10135988  
; Publication No. US20020176869A1  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175
US-10-135-988-4

Query Match      100.0%; Score 102; DB 13; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PYKELRLEVGKQRLKYAQEE 20
Db      799 PYKELRLEVGKQRLKYAQEE 818

RESULT 9
US-10-135-988-6
; Sequence 6, Application US/10135988
; Publication No. US20020176869A1
; GENERAL INFORMATION:
; APPLICANT: FUNKHOUSER, ANN W
; APPLICANT: EMERSON, SUZANNE U
; APPLICANT: PURCELL, ROBERT H
; APPLICANT: D'HONDT, ERIC
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES
; FILE REFERENCE: 20264262US3
; CURRENT APPLICATION NUMBER: US/10/135,988
; CURRENT FILING DATE: 2002-04-29
; PRIOR FILING DATE: 07/947,338
; PRIOR FILING DATE: 1992-09-18
; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-10-135-988-6

Query Match      100.0%; Score 102; DB 13; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PYKELRLEVGKQRLKYAQEE 20
Db      799 PYKELRLEVGKQRLKYAQEE 818

RESULT 10
US-09-171-432A-39
; Sequence 39, Application US/09171432A
; Publication No. US20030187184A1
; GENERAL INFORMATION:
; APPLICANT: Fields, Howard A.
; APPLICANT: Khudyakov, Yury E.
; TITLE OF INVENTION: Antigenically Reactive Regions of the
; TITLE OF INVENTION: Hepatitis A Virus Polypeptide
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick Stockton LLP
; STREET: 3424 Peachtree Road, N.E.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30326
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/09/171,432A
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Greene, Jamie L.
; REGISTRATION NUMBER: 32,467
; REFERENCE/DOCKET NUMBER: 03063-0231US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 949-2400
; TELEFAX: (404) 949-2499
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..20
; OTHER INFORMATION: /label= YK-1315
US-09-171-432A-39

Query Match      65.7%; Score 67; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00064;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PYKELRLEVGKQR 13
Db      8 PYKELRLEVGKQR 20

RESULT 11
US-09-864-408A-140
; Sequence 140, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encc
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-140

Query Match      45.6%; Score 46.5; DB 11; Length 65;
Best Local Similarity 55.6%; Pred. No. 4.9;
Matches 10; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY      1 PYKELRLEVGKQRLKYAQ 18
Db      34 PYKGLR-EISRNMLEFYAQ 50

RESULT 12
US-09-171-432A-41
; Sequence 41, Application US/09171432A
; Publication No. US20030187184A1
; GENERAL INFORMATION:
; APPLICANT: Fields, Howard A.
; APPLICANT: Khudyakov, Yury E.
; TITLE OF INVENTION: Antigenically Reactive Regions of the
```

; TITLE OF INVENTION: Hepatitis A Virus Polypeptide  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick Stockton LLP  
; STREET: 3424 Peachtree Road, N.E.  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30326  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/171,432A  
; FILING DATE: 23-NOV-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Greene, Jamie L.  
; REGISTRATION NUMBER: 32,467  
; REFERENCE/DOCKET NUMBER: 03063-0231US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 949-2400  
; TELEFAX: (404) 949-2499  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..20  
; OTHER INFORMATION: /label=YK-1317  
; US-09-171-432A-41  
  
Query Match 44.1%; Score 45; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 12 QRLKYAQEE 20  
DB 1 QRLKYAQEE 9  
  
RESULT 13  
US-09-171-432A-47  
; Sequence 47, Application US/09171432A  
; Publication No. US20030187184A1  
; GENERAL INFORMATION:  
; APPLICANT: Fields, Howard A.  
; APPLICANT: Khudiyakov, Yuri E.  
; TITLE OF INVENTION: Antigenically Reactive Regions of the  
; TITLE OF INVENTION: Hepatitis A Virus Polypeptide  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick Stockton LLP  
; STREET: 3424 Peachtree Road, N.E.  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30326  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/171,432A  
; FILING DATE: 23-NOV-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Greene, Jamie L.  
; REGISTRATION NUMBER: 32,467  
; REFERENCE/DOCKET NUMBER: 03063-0231US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 949-2400  
; TELEFAX: (404) 949-2499  
; INFORMATION FOR SEQ ID NO: 47:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..25  
; OTHER INFORMATION: /label=YK-1665  
; US-09-171-432A-47  
  
Query Match 44.1%; Score 45; DB 10; Length 25;  
Best Local Similarity 100.0%; Pred. No. 3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 12 QRLKYAQEE 20  
DB 1 QRLKYAQEE 9  
  
RESULT 14  
US-10-289-762-108  
; Sequence 108, Application US/10289762  
; Publication No. US20040006218A1  
; GENERAL INFORMATION:  
; APPLICANT: Griffais, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/10/289,762  
; CURRENT FILING DATE: 2003-03-27  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 108  
; LENGTH: 214  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
; US-10-289-762-108  
  
Query Match 44.1%; Score 45; DB 15; Length 214;  
Best Local Similarity 50.0%; Pred. No. 32;  
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
  
QY 3 KELRLEVGKORLYAQEE 20  
DB 138 KYMAEVSKERLKLSE 155  
  
RESULT 15  
US-10-238-075-1284  
; Sequence 1284, Application US/10238075  
; Publication No. US20030148324A1  
; GENERAL INFORMATION:  
; APPLICANT: I.N.S.E.R.M.  
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated  
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their  
; FILE REFERENCE: BLANDINE  
; CURRENT APPLICATION NUMBER: US/10/238,075

US-10-312-373-1

Query Match 43.1%; Score 44; DB 14; Length 211;  
Best Local Similarity 42.9%; Pred. No. 45;  
Matches 9; Conservative 7; Mismatches 3; Indels 2; Gaps 1;

QY 1 PYKELRLEVGKQRLKYAQE 19  
|||:|||||:  
Db 135 PYQIAREVKVPRRLKYALE 155

RESULT 16

US-10-424-599-252284  
; Sequence 252284, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 252284  
; LENGTH: 270  
; TYPE: PRT  
; ORGANISM: Glycine max

FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_69840C.1.p.1ep

US-10-424-599-252284

Query Match 43.1%; Score 44; DB 12; Length 270;  
Best Local Similarity 52.9%; Pred. No. 60;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 YKELRLEVGKQRLKYAQ 18  
|||:|||||:  
Db 22 KYFLIRVGKQRLREAR 38

RESULT 17

US-10-312-373-1

; Sequence 1, Application US/10312373  
; Publication No. US20040072174A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehm, Thomas  
; APPLICANT: Dear, Neil T  
; TITLE OF INVENTION: CALPAIN PROTEASE 12  
; FILE REFERENCE: RKG-001US  
; CURRENT APPLICATION NUMBER: US/10/312,373  
; CURRENT FILING DATE: 2002-12-23  
; PRIOR APPLICATION NUMBER: PCT/EP01/07457  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: DE 10031932.7  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 720  
; TYPE: PRT  
; ORGANISM: Mus sp.

US-10-312-373-1

Query Match 43.1%; Score 44; DB 12; Length 720;  
Best Local Similarity 50.0%; Pred. No. 1.8e+02;  
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
|||:|||||:  
Db 545 PYKPLELELAQLFLELAGEE 564

RESULT 18

US-10-390-585-16  
; Sequence 16, Application US/10390585  
; Publication No. US20040014093A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL CYSTEINE PROTEASE OF THE CALPAIN  
; FILE REFERENCE: D0219DP  
; CURRENT APPLICATION NUMBER: US/10/390,585  
; CURRENT FILING DATE: 2003-03-14  
; PRIOR APPLICATION NUMBER: U.S. 60/364,941  
; PRIOR FILING DATE: 2002-03-14  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 16  
; LENGTH: 720  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-390-585-16

Query Match 43.1%; Score 44; DB 15; Length 720;  
Best Local Similarity 50.0%; Pred. No. 1.8e+02;  
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 PYKELRLEVGKQRLKYAQEE 20  
|||:|||||:  
Db 545 PYKPLELELAQLFLELAGEE 564

RESULT 19

US-10-055-569A-35  
; Sequence 35, Application US/10055569A  
; Publication No. US20040024181A1  
; GENERAL INFORMATION:  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Gilbert, Jennifer  
; APPLICANT: Casman, Stacie  
; APPLICANT: Blalock, Angela  
; APPLICANT: Li, Li  
; APPLICANT: Vernet, Corine  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Mishra, Vishnu S  
; APPLICANT: Furtak, Katarzyna  
; APPLICANT: Gerlach, Valerie L  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Malyanker, Uriel  
; APPLICANT: Stone, David  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Gunther, Erik  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Taupier Jr., Raymond J  
; APPLICANT: Anderson, David W  
; TITLE OF INVENTION: No. US20040024181A1 Human Proteins, Polynucleotides Encoding Th  
; FILE REFERENCE: 21402-191  
; CURRENT APPLICATION NUMBER: US/10/055,569A  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: 60/243,642

;; PRIOR FILING DATE: 2000-10-26  
;; PRIOR APPLICATION NUMBER: 60/243,320  
;; PRIOR FILING DATE: 2000-10-26  
;; PRIOR APPLICATION NUMBER: 60/243,592  
;; PRIOR FILING DATE: 2000-10-26  
;; PRIOR APPLICATION NUMBER: 60/243,681  
;; PRIOR FILING DATE: 2000-10-27  
;; PRIOR APPLICATION NUMBER: 60/243,863  
;; PRIOR FILING DATE: 2000-10-27  
;; PRIOR APPLICATION NUMBER: 60/244,443  
;; PRIOR FILING DATE: 2000-10-31  
;; PRIOR APPLICATION NUMBER: 60/245,029  
;; PRIOR FILING DATE: 2000-11-01  
;; PRIOR APPLICATION NUMBER: 60/244,995  
;; PRIOR FILING DATE: 2000-10-27  
;; PRIOR APPLICATION NUMBER: 60/244,443  
;; PRIOR FILING DATE: 2000-10-31  
;; PRIOR APPLICATION NUMBER: 60/245,029  
;; PRIOR FILING DATE: 2000-11-01  
;; PRIOR APPLICATION NUMBER: 60/244,995  
;; PRIOR FILING DATE: 2000-11-01  
;; PRIOR APPLICATION NUMBER: 60/245,293  
;; PRIOR FILING DATE: 2000-11-02  
;; PRIOR APPLICATION NUMBER: 60/245,315  
;; PRIOR FILING DATE: 2000-11-02  
;; PRIOR APPLICATION NUMBER: 60/245,315  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 137  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 35  
;; LENGTH: 720  
;; TYPE: PRT  
;; ORGANISM: Mus musculus  
US-10-055-569A-35

Query Match 43.1%; Score 44; DB 16; Length 720;  
Best Local Similarity 50.0%; Pred. No. 1.8e+02;  
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 PYKLELEVGKQRLKYAOEE 20  
DB 545 PYKLELELAQLFLELAGEE 564

RESULT 20  
US-10-055-569A-36  
; Sequence 36, Application US/10055569A  
; Publication No. US20040024181A1  
; GENERAL INFORMATION:  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Gilbert, Jennifer  
; APPLICANT: Casman, Stacie  
; APPLICANT: Blalock, Angela  
; APPLICANT: Li, Li  
; APPLICANT: Vernet, Corine  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Mishra, Vishnu S  
; APPLICANT: Furtak, Katarzyna  
; APPLICANT: Gerlach, Valerie L  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Malyanker, Uriel  
; APPLICANT: Stone, David  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Gunther, Erik  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Taupier Jr., Raymond J  
; APPLICANT: Anderson, David W  
; TITLE OF INVENTION: No. US20040024181A1 Human Proteins, Polynucleotides Encoding Th  
; TITLE OF INVENTION: Methods of Using the Same  
; FILE REFERENCE: 21402-191  
; CURRENT APPLICATION NUMBER: US/10/055,569A  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: 60/243,642  
; PRIOR FILING DATE: 2000-10-26  
; PRIOR APPLICATION NUMBER: 60/243,320  
; PRIOR FILING DATE: 2000-10-26  
; PRIOR APPLICATION NUMBER: 60/243,592  
; PRIOR FILING DATE: 2000-10-26  
; PRIOR APPLICATION NUMBER: 60/243,320  
; PRIOR FILING DATE: 2000-10-26  
; PRIOR APPLICATION NUMBER: 60/243,592

;; PRIOR FILING DATE: 2000-10-26  
;; PRIOR APPLICATION NUMBER: 60/243,681  
;; PRIOR FILING DATE: 2000-10-27  
;; PRIOR APPLICATION NUMBER: 60/243,863  
;; PRIOR FILING DATE: 2000-10-27  
;; PRIOR APPLICATION NUMBER: 60/244,443  
;; PRIOR FILING DATE: 2000-10-31  
;; PRIOR APPLICATION NUMBER: 60/245,029  
;; PRIOR FILING DATE: 2000-11-01  
;; PRIOR APPLICATION NUMBER: 60/244,995  
;; PRIOR FILING DATE: 2000-11-01  
;; PRIOR APPLICATION NUMBER: 60/245,293  
;; PRIOR FILING DATE: 2000-11-02  
;; PRIOR APPLICATION NUMBER: 60/245,315  
;; PRIOR FILING DATE: 2000-11-02  
;; PRIOR APPLICATION NUMBER: 60/245,315  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 137  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 36  
;; LENGTH: 720  
;; TYPE: PRT  
;; ORGANISM: Mus musculus  
US-10-055-569A-36

Query Match 43.1%; Score 44; DB 16; Length 720;  
Best Local Similarity 50.0%; Pred. No. 1.8e+02;  
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 PYKLELEVGKQRLKYAOEE 20  
DB 545 PYKLELELAQLFLELAGEE 564

RESULT 21  
US-10-055-569A-37  
; Sequence 37, Application US/10055569A  
; Publication No. US20040024181A1  
; GENERAL INFORMATION:  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Gilbert, Jennifer  
; APPLICANT: Casman, Stacie  
; APPLICANT: Blalock, Angela  
; APPLICANT: Li, Li  
; APPLICANT: Vernet, Corine  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Mishra, Vishnu S  
; APPLICANT: Furtak, Katarzyna  
; APPLICANT: Gerlach, Valerie L  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Malyanker, Uriel  
; APPLICANT: Stone, David  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Gunther, Erik  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Taupier Jr., Raymond J  
; APPLICANT: Anderson, David W  
; TITLE OF INVENTION: No. US20040024181A1 Human Proteins, Polynucleotides Encoding The  
; TITLE OF INVENTION: Methods of Using the Same  
; FILE REFERENCE: 21402-191  
; CURRENT APPLICATION NUMBER: US/10/055,569A  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: 60/243,642  
; PRIOR FILING DATE: 2000-10-26  
; PRIOR APPLICATION NUMBER: 60/243,320  
; PRIOR FILING DATE: 2000-10-26  
; PRIOR APPLICATION NUMBER: 60/243,592  
; PRIOR FILING DATE: 2000-10-26  
; PRIOR APPLICATION NUMBER: 60/243,681  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: 60/243,863

; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: 60/244,443
; PRIOR FILING DATE: 2000-10-31
; PRIOR APPLICATION NUMBER: 60/245,029
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/244,995
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/245,293
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 60/245,315
; PRIOR FILING DATE: 2000-11-02
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-055-569A-37

Query Match 43.1%; Score 44; DB 16; Length 720;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PYKELRELVGKQRLKYAQEE 20
Db 545 PYKPLELELAQLELAE 564

RESULT 22
US-10-363-616-325
; Sequence 325, Application US/10363616
; Publication No. US20040044181A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-113 (793)
; CURRENT APPLICATION NUMBER: US/10/363,616
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: 09/654,935
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 490
; SEQ ID NO 325
; LENGTH: 1273
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-363-616-325

Query Match 42.6%; Score 43.5; DB 12; Length 1273;
Best Local Similarity 55.6%; Pred. No. 4e+02;
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Qy 3 KELRELVGKQRL-KYAE 19
Db 109 KELRMEIAKQELIAHARE 126

RESULT 23
US-10-238-075-999
; Sequence 999, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 999

; LENGTH: 120
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (116)..(116)
; OTHER INFORMATION: unknown
; US-10-238-075-999

Query Match 42.2%; Score 43; DB 14; Length 120;
Best Local Similarity 47.6%; Pred. No. 35;
Matches 10; Conservative 5; Mismatches 4; Indels 2; Gaps 1;

Qy 1 PYKELRELV--GKORLYAQEE 19
Db 67 PYQVAREVKAPRERLYALE 87

RESULT 24
US-10-424-599-265976
; Sequence 265976, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 265976
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_82197C.1.pep
; US-10-424-599-265976

Query Match 42.2%; Score 43; DB 12; Length 136;
Best Local Similarity 45.0%; Pred. No. 41;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 1 PYKELRELVGKQRLKYAQEE 20
Db 75 PYKLGDDGVDGLHYCPDE 94

RESULT 25
US-10-091-007-124
; Sequence 124, Application US/10091007
; Publication No. US20030170782A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics limited
; APPLICANT: Le Page, Richard W F
; APPLICANT: Hannify, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21978WO
; CURRENT APPLICATION NUMBER: US/10/091,007
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: GB 9921125.2
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 124
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
; US-10-091-007-124

Query Match 42.2%; Score 43; DB 14; Length 178;



Best Local Similarity 58.3%; Pred. No. 55;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 PYKELRLVGVKQ 12  
DB 117 PFKELRLSAGRE 128

RESULT 26

US-10-424-599-216526  
; Sequence 216526, Application US/10424599  
; Publication No. US20040031072A1

GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 216526

; LENGTH: 90

; TYPE: PRT

; ORGANISM: Glycine max

FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_37551C.1.pep

US-10-424-599-216526

Query Match 41.2%; Score 42; DB 12; Length 90;

Best Local Similarity 42.1%; Pred. No. 38;

Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 PYKELRLVGVKQRLKYAQE 19

DB 48 PERSIGLPLKHLRYVQE 66

RESULT 27

US-10-424-599-226551  
; Sequence 226551, Application US/10424599  
; Publication No. US20040031072A1

GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 226551

; LENGTH: 156

; TYPE: PRT

; ORGANISM: Glycine max

FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)-(156)

; OTHER INFORMATION: unsure at all Xaa locations

FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_46604C.1.pep

US-10-424-599-226551

Query Match 41.2%; Score 42; DB 12; Length 156;

Best Local Similarity 52.9%; Pred. No. 69;

Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 YKELRLVGVKQRLKYAQ 18

DB 117 PFKELRLSAGRE 128

DB 69 YGELRPPLGKHLKIVE 85

RESULT 28

US-10-238-075-1123  
; Sequence 1123, Application US/10238075  
; Publication No. US20030148324A1

GENERAL INFORMATION:

; APPLICANT: I.N.S.E.R.M.

; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolate  
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their

; FILE REFERENCE: BLANDINE

; CURRENT APPLICATION NUMBER: US/10/238,075

; CURRENT FILING DATE: 2002-09-10

; PRIOR APPLICATION NUMBER: 0003145

; PRIOR FILING DATE: 2000-03-10

; NUMBER OF SEQ ID NOS: 1576

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 1123

; LENGTH: 207

; TYPE: PRT

; ORGANISM: Escherichia coli

US-10-238-075-1123

Query Match 41.2%; Score 42; DB 14; Length 207;

Best Local Similarity 52.4%; Pred. No. 94;

Matches 11; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 1 PYKELRLVGVKQRLKYAQE 19

DB 131 PYQLIAREVKAPQRLKYALE 151

RESULT 29

US-10-424-599-212457  
; Sequence 212457, Application US/10424599  
; Publication No. US20040031072A1

GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 212457

; LENGTH: 443

; TYPE: PRT

; ORGANISM: Glycine max

FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_33874C.1.pep

US-10-424-599-212457

Query Match 41.2%; Score 42; DB 12; Length 443;

Best Local Similarity 53.3%; Pred. No. 2.2e+02;

Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 6 RLEVGVKQRLKYAQEE 20

DB 112 RLNVGKEQVQAQNE 126

RESULT 30

US-10-282-122A-76680  
; Sequence 76680, Application US/10282122A  
; Publication No. US20040029129A1

GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76680
; LENGTH: 736
; TYPE: PRT
; ORGANISM: Ureaplasma urealyticum
US-10-282-122A-76680

Query Match      41.2%; Score 42; DB 12; Length 736;
Best Local Similarity 44.4%; Pred. No. 3.8e+02;
Matches      8; Conservative      4; Mismatches      6; Indels      0; Gaps      0;

QY      2 YKELELVGKQRLKYAQE 19
DB      396 YKAIYSEINKKKNQYLQE 413

Search completed: May 11, 2004, 13:44:49
Job time : 35.25 secs

```

GenCore version 5.1.6 Copyright (c) 1993 - 2004 CompuGen Ltd.																										
OM protein - protein search, using sw model																										
Run on:	May 11, 2004, 13:23:11	Search time 47 Seconds (without alignments) 120.233 Million cell updates/sec																								
Title:	US-09-171-432A-41																									
Perfect score:	103																									
Sequence:	1 QRLKYAOEELSNEVLPPPRK 20																									
Scoring table:	BLOSUM62 Gapop 10.0 , Gapext 0.5																									
Searched:	1586107 seqs, 282547505 residues																									
Total number of hits satisfying chosen parameters:	1586107																									
Minimum DB seq length:	0																									
Maximum DB seq length:	2000000000																									
Post-processing:	Minimum Match 0% Maximum Match 100% Listing first 100 summaries																									
Database :	A_Geneseq_29Jan04.* 1: Geneseqp1980s.* 2: Geneseqp1990s.* 3: Geneseqp2000s.* 4: Geneseqp2001s.* 5: Geneseqp2002s.* 6: Geneseqp2003as.* 7: Geneseqp2003bs.* 8: Geneseqp2004s.*																									
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.																										
SUMMARIES																										
Result No.	Score	Query Match	Length	DB ID	Description																					
1	103	100.0	20	2	Aaw42924 Immunogen																					
2	103	100.0	21	4	Aab69441 Synthetic																					
3	103	100.0	25	2	Aaw42930 Immunogen																					
4	103	100.0	26	4	Aab69447 Synthetic																					
5	103	100.0	366	1	Aap50230 Sequence																					
6	103	100.0	854	1	Aap50287 Hepatitis																					
7	103	100.0	993	1	Aap50116 Sequence																					
8	103	100.0	993	1	Aap50231 Sequence																					
9	103	100.0	1077	2	Aaw95559 A partial																					
10	103	100.0	1091	2	Aar32426 Translate																					
11	103	100.0	2227	1	Aap60066 Sequence																					
12	103	100.0	2227	2	Aar05697 Attenuate																					
13	103	100.0	2227	2	Aaw34074 Hepatitis																					
14	103	100.0	2227	3	Aab18609 Amino aci																					
15	103	100.0	2227	3	Aab18607 Amino aci																					
16	103	100.0	2227	5	Aab18999 Hepatitis																					
17	103	100.0	2227	5	Aab311729 Attenuate																					
18	103	100.0	2227	5	Aab31172 Wild-type																					
19	103	100.0	2227	6	Aab08641 Attenuat																					
20	103	100.0	2227	6	Aab08639 Wild type																					
21	103	100.0	2227	7	Aab00350 Hepatitis																					
22	98	95.1	2227	3	Aab18608 Amino aci																					
23	98	95.1	2227	5	Aab311728 Hepatitis																					
24	98	95.1	2227	6	Aab08640 Attenuat																					
25	94	91.3	839	2	Aar15629 Capsid re																					

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:23:11 ; Search time 47 Seconds  
(without alignments)  
120.233 Million cell updates/sec

Title: US-09-171-432A-41

Perfect score: 103

Sequence: 1 QRLKYAOEELSNEVLPPPRK 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	100.0	20	2	Aaw42924 Immunogen
2	103	100.0	21	4	Aab69441 Synthetic
3	103	100.0	25	2	Aaw42930 Immunogen
4	103	100.0	26	4	Aab69447 Synthetic
5	103	100.0	366	1	Aap50230 Sequence
6	103	100.0	854	1	Aap50287 Hepatitis
7	103	100.0	993	1	Aap50116 Sequence
8	103	100.0	993	1	Aap50231 Sequence
9	103	100.0	1077	2	Aaw95559 A partial
10	103	100.0	1091	2	Aar32426 Translate
11	103	100.0	2227	1	Aap60066 Sequence
12	103	100.0	2227	2	Aar05697 Attenuate
13	103	100.0	2227	2	Aaw34074 Hepatitis
14	103	100.0	2227	3	Aab18609 Amino aci
15	103	100.0	2227	3	Aab18607 Amino aci
16	103	100.0	2227	5	Aab18999 Hepatitis
17	103	100.0	2227	5	Aab311729 Attenuate
18	103	100.0	2227	5	Aab31172 Wild-type
19	103	100.0	2227	6	Aab08641 Attenuat
20	103	100.0	2227	6	Aab08639 Wild type
21	103	100.0	2227	7	Aab00350 Hepatitis
22	98	95.1	2227	3	Aab18608 Amino aci
23	98	95.1	2227	5	Aab311728 Hepatitis
24	98	95.1	2227	6	Aab08640 Attenuat
25	94	91.3	839	2	Aar15629 Capsid re

99 42.5 41.3 530 3 AAB42529 Human ORF  
100 42.5 41.3 530 5 ABP68940 Human pol

## ALIGNMENTS

## RESULT 1

AAW42924  
ID AAW42924 standard; peptide; 20 AA.

XX AC AAW42924;

XX DT 28-APR-1998 (first entry)

XX DE Immunogenic Hepatitis A virus peptide YK-1317.

XX KW Immunogenic peptide; immunogenic epitope; P2A protein; immune response;

XX OS Synthetic.

XX OS Hepatitis A virus.

XX PN WO9740147-A1.

XX PD 30-OCT-1997.

XX PF 18-APR-1997; 97WO-US006891.

XX PR 19-APR-1996; 96US-0015644P.

XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX PI Fields HA, Khudyakov YE;

XX DR WPI; 1997-535831/49.

XX PT Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune

XX PT response to HAV in a mammal or to detect the presence of antibodies

XX PT against HAV in a mammal.

XX PS Claim 18; Page 112; 140pp; English.

XX CC Peptides AAW42922-30 are immunogenic peptides corresponding to

XX CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are

XX CC substantially similar to a portion of the amino acid sequence of the P2A

XX CC protein of HAV corresponding to amino acids 792-980. The present peptide

XX CC is derived from amino acids 810-829, and has a reactivity of 83.3% with

XX CC acute sera. Compositions containing the peptides can be used to induce an

XX CC immune response to HAV in a mammal. The peptides can also be used to

XX CC detect the presence of antibodies against HAV in mammalian serum. The

XX CC peptides can also be used to make an antibody against HAV by

XX CC administering the peptide to a mammal

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 103; DB 2; Length 20;

Best Local Similarity 100.0%; Pred. No. 8.9e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20

DB 1 QRLKYAQEELSNEVLPPPRK 20

RESULT 2

AAB69441

ID AAB69441 standard; peptide; 21 AA.

XX AC AAB69441;

XX DT 20-APR-2001 (first entry)

XX PN WO9740147-A1.

XX

DE Synthetic HAV P2A peptide, SEQ ID NO: 41.

XX Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;

KW antigen; major structural capsid polypeptide; HAV antibody detection.

XX Hepatitis A virus.

OS Synthetic.

XX WO200105824-A2.

XX PD 25-JAN-2001.

XX PF 14-JUL-2000; 2000WO-US019267.

XX PR 15-JUL-1999; 99US-0144412P.

XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX PI Fields HA, Khudyakov YE;

XX DR WPI; 2001-112681/12.

XX Synthetic peptides used as antigen sources for enzyme immunoassays

PT detecting anti-hepatitis A virus and as vaccines.

XX Claim 13; Page 94; 130pp; English.

XX The present sequence is one of a number of synthetic peptides which are

XX immunoreactive with hepatitis A virus (HAV) antibodies. The peptides

XX comprise antigenic epitopes of the major structural capsid polypeptides

XX or non-structural polypeptides of HAV with one or more glutamine

XX molecules at the carboxy end of the peptide. The peptides are used to

XX detect the presence of antibodies against HAV in mammalian serum, to

XX detect the presence of HAV in a human or animal through the binding of

XX the peptide to an antibody, to detect acute phase infection by detecting

XX IGM antibodies in mammalian serum and detecting convalescence in a

XX mammal. The peptides are used to detect or quantify HAV antibodies in

XX samples in clinical or research-based assays using immunoblotting.

XX fluorescence in situ hybridisation analysis, gel-mobility shift assays,

XX tracking of radioactive or bioluminescent markers, chromatography or

XX electrophoresis. The peptides are used to induce an immune response to

XX HAV when administered to a human or animal. Glutamine at the carboxy end

XX of the peptides enhances the IGM antibody reactivity

XX SQ Sequence 21 AA;

Query Match 100.0%; Score 103; DB 4; Length 21;

Best Local Similarity 100.0%; Pred. No. 9.4e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20

DB 1 QRLKYAQEELSNEVLPPPRK 20

RESULT 3

AAW42930

ID AAW42930 standard; peptide; 25 AA.

XX AC AAW42930;

XX DT 28-APR-1998 (first entry)

XX DE Immunogenic Hepatitis A virus peptide YK-1665.

XX Immunogenic peptide; immunogenic epitope; P2A protein; immune response;

XX KW antibody.

XX OS Synthetic.

XX OS Hepatitis A virus.

XX PN WO9740147-A1.

XX

PD 30-OCT-1997.  
 XX 18-APR-1997; 97WO-US006891.  
 XX 19-APR-1996; 96US-0015644P.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Fields HA, Khudyakov YE;  
 XX WPI; 1997-535831/49.  
 XX Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
 PT response to HAV in a mammal or to detect the presence of antibodies  
 PT against HAV in a mammal.  
 XX Claim 18; Page 112; 140pp; English.  
 XX Peptides AAW42922-30 are immunogenic peptides corresponding to  
 CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
 CC substantially similar to a portion of the amino acid sequence of the P2A  
 CC protein of HAV corresponding to amino acids 792-980. Compositions  
 CC containing the peptides can be used to induce an immune response to HAV  
 CC in a mammal. The peptides can also be used to detect the presence of  
 CC antibodies against HAV in mammalian serum. The peptides can also be used  
 CC to make an antibody against HAV by administering the peptide to a mammal  
 XX  
 SQ Sequence 25 AA;  
 Query Match 100.0%; Score 103; DB 2; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 DB 1 QRLKYAQEELSNEVLPPPRK 20  
 RESULT 4  
 AAB69447  
 ID AAB69447 standard; peptide; 26 AA.  
 XX AC AAB69447;  
 XX 20-APR-2001 (first entry)  
 XX Synthetic HAV P2A peptide, SEQ ID NO: 47.  
 DE Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
 XX antigen; major structural capsid polypeptide; HAV antibody detection.  
 KW Hepatitis A virus.  
 XX Synthetic.  
 OS WO200105824-A2.  
 PN 25-JAN-2001.  
 XX 14-JUL-2000; 2000WO-US019267.  
 PF 15-JUL-1999; 99US-0144412P.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Fields HA, Khudyakov YE;  
 XX WPI; 2001-112681/12.  
 XX Synthetic peptides used as antigen sources for enzyme immunoassays  
 PT detecting anti-hepatitis A virus and as vaccines.  
 XX Claim 13; Page 98; 130pp; English.

CC The present sequence is one of a number of synthetic peptides which are  
 CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
 CC comprise antigenic epitopes of the major structural capsid polypeptides  
 CC or non-structural polypeptides of HAV with one or more glutamine  
 CC molecules at the carboxy end of the peptide. The peptides are used to  
 CC detect the presence of antibodies against HAV in mammalian serum, to  
 CC detect the presence of HAV in a human or animal through the binding of  
 CC the peptide to an antibody, to detect acute phase infection by detecting  
 CC IGM antibodies in mammalian serum and detecting convalescence in a  
 CC mammal. The peptides are used to detect or quantify HAV antibodies in  
 CC samples in clinical or research-based assays using immunoblotting,  
 CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
 CC tracking of radioactive or bioluminescent markers, chromatography or  
 CC electrophoresis. The peptides are used to induce an immune response to  
 CC HAV when administered to a human or animal. Glutamine at the carboxy end  
 CC of the peptides enhances the IGM antibody reactivity  
 XX  
 SQ Sequence 26 AA;  
 Query Match 100.0%; Score 103; DB 4; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 DB 1 QRLKYAQEELSNEVLPPPRK 20  
 RESULT 5  
 AAP50230  
 ID AAP50230 standard; protein; 366 AA.  
 XX AC AAP50230;  
 XX 28-NOV-1991 (first entry)  
 XX Sequence of hepatitis A virus (HAV) surface protein (VP-1).  
 XX Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
 KW diagnostic assay.  
 XX Hepatitis A virus.  
 OS EP138704-A.  
 PN 24-APR-1985.  
 XX 09-OCT-1984; 84EP-00402025.  
 XX 14-OCT-1983; 83US-00541836.  
 PR 02-MAR-1984; 84US-00585942.  
 XX (MERI ) MERCK & CO INC.  
 XX Hughes JV, Scolnick EM, Tomassini JE;  
 XX WPI; 1985-100818/17.  
 DR N-PSDB; AAN50274.  
 XX New hepatitis A virus surface protein - useful for binding to  
 PT neutralising antibodies to the virus.  
 XX Claim 21; Page 46-48; 49pp; English.  
 XX VP1 is isolated by solubilisation of the intact virus in an aq. anionic  
 CC surfactant and a reducing agent. The viral proteins are sepd. and the  
 CC protein of molecular wt. 33000 daltons is sepd  
 XX  
 SQ Sequence 366 AA;  
 Query Match 100.0%; Score 103; DB 1; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 319 QRLKYAQEELSNEVLPPPRK 338

RESULT 6  
AAP50287  
ID AAP50287 standard; protein; 854 AA.  
XX AAP50287;  
AC  
XX 25-MAR-2003 (revised)  
DT 30-NOV-1991 (first entry)  
XX  
XX Hepatitis A virus (HAV) peptide corresponding to the capsid protein  
DE region of poliovirus RNA.  
DE  
XX  
XX Hepatitis A virus assay; antigen; antibody.  
XX  
XX Hepatitis A virus.  
OS  
XX  
XX WO8501517-A.  
XX  
XX 11-APR-1985.  
XX  
XX 27-SEP-1984; 84WO-US001552.  
XX  
XX 30-SEP-1983; 83US-00537911.  
XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX  
XX Ticehurst JR, Baltimore D, Feinstein SM, Purcell RH;  
PI Racaniello VR;  
FI  
XX WPI; 1985-098846/16.  
DR N-PSDB; AAN50330.  
XX  
XX New hepatitis A virus CDNA - useful in assays for the virus and for  
PT prodn. of the viral antigen and antibodies to it.  
XX  
XX Example; Fig 7; 60pp; English.  
FS  
XX The inventors claim HAV cDNA and a method for producing it, whereby large  
CC ants. can be obtd. economically. The cDNA is useful in the assay for  
CC detection of HAV quickly and easily and with high sensitivity and  
CC specificity. The HAV cDNA is also used in the prodn. of HAV antigen or  
CC antibodies to it. The antibodies may be monoclonal. (Updated on 25-MAR-  
CC 2003 to correct PA field.)  
XX  
XX Sequence 854 AA;  
SQ

Query Match 100.0%; Score 103; DB 1; Length 854;  
Best Local Similarity 100.0%; Pred. No. 4.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 7  
AAP50116  
ID AAP50116 standard; protein; 993 AA.  
XX AAP50116;  
AC  
XX 24-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 30-SEP-1991 (first entry)  
XX  
XX Sequence of Hepatitis A virus (HAV) immunogenic peptides VP-1, VP-2, VP-3  
DE and VP-4.

Query Match 100.0%; Score 103; DB 1; Length 854;  
Best Local Similarity 100.0%; Pred. No. 4.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 7  
AAP50116  
ID AAP50116 standard; protein; 993 AA.  
XX AAP50116;  
AC  
XX 24-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 30-SEP-1991 (first entry)  
XX  
XX Sequence of Hepatitis A virus (HAV) immunogenic peptides VP-1, VP-2, VP-3  
DE and VP-4.

Query Match 100.0%; Score 103; DB 1; Length 993;  
Best Local Similarity 100.0%; Pred. No. 5.3e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 946 QRLKYAQEELSNEVLPPPRK 965

RESULT 8  
AAP50231  
ID AAP50231 standard; protein; 993 AA.  
XX AAP50231;  
AC  
XX 28-NOV-1991 (first entry)  
DT  
XX  
DE Sequence encoded by partial sequence of hepatitis A virus (HAV),  
DE including surface protein (VP-1).  
XX  
XX Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
XX diagnostic assay.  
XX  
XX Hepatitis A virus.  
OS  
XX  
XX Key Location/Qualifiers  
FT 628..993  
FT /note="claimed; X denotes translated stop codons and  
FT unspecified triplets"  
XX  
XX EP138704-A.  
XX  
XX 24-APR-1985.  
PD

XX Antigenic protein; immunogen; vaccine.  
XX  
XX Hepatitis A virus; (strain CR326).  
OS  
XX  
XX EP154587-A.  
PN  
XX  
XX 11-SEP-1985.  
PD  
XX  
XX 27-FEB-1985; 85EP-00400369.  
PF  
XX  
XX 02-MAR-1984; 84US-00585818.  
PR  
XX (MERI ) MERCK & CO INC.  
XX  
XX Linemeyer DL, Menke JG, Rueben RG, Mitra SW;  
PI  
XX WPI; 1985-224964/37.  
DR N-PSDB; AAN50139.  
DR  
XX  
XX New nucleotide sequences coding for hepatitis A virus antigens - useful  
PT for eliciting normal immune response and in vaccines for protecting  
PT against the virus.  
XX  
XX Example; Page 11-17; 32pp; English.  
PS  
XX  
XX Within the sequence in AAN50139 is encoded the information necessary to  
CC make the antigenic proteins of HAV. The sequences encoding for the  
CC structural proteins begin at base 403. The key sub-unit sequences within  
CC VP-1, designated Sequences I,II,III,IV, and V, start, respectively at  
CC 1882, 1963, 1999, 2146, 2347. Other nucleotide sequences which are  
CC valuable as encoding antigenic proteins are the sequences from base 1749  
CC to base 2722; from base 1487 to base 2980 and from base 1644 to base  
CC 2722. The sequence from base 1749 to base 2722 is esp. valuable as a  
CC vector for producing antigen protein. Sequences II-V are claimed. X in  
CC AAP50116 denotes the translation of a stop codon. (Updated on 25-MAR-2003  
CC to correct PI field.) (Updated on 24-OCT-2003 to standardise OS field)  
XX  
XX Sequence 993 AA;  
SQ

Query Match 100.0%; Score 103; DB 1; Length 993;  
Best Local Similarity 100.0%; Pred. No. 5.3e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 946 QRLKYAQEELSNEVLPPPRK 965

RESULT 8  
AAP50231  
ID AAP50231 standard; protein; 993 AA.  
XX AAP50231;  
AC  
XX 28-NOV-1991 (first entry)  
DT  
XX  
DE Sequence encoded by partial sequence of hepatitis A virus (HAV),  
DE including surface protein (VP-1).  
XX  
XX Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
XX diagnostic assay.  
XX  
XX Hepatitis A virus.  
OS  
XX  
XX Key Location/Qualifiers  
FT 628..993  
FT /note="claimed; X denotes translated stop codons and  
FT unspecified triplets"  
XX  
XX EP138704-A.  
XX  
XX 24-APR-1985.  
PD

XX -type HAV genome in which the P2 region is replaced by the P2 region from  
 CC a cell-culture-adapted HAV strain. The construct is used to demonstrate  
 CC that mutations in the P2 region of a cell-culture-adapted HAV strain are  
 CC sufficient for establishment of infection and accelerated growth in cell  
 CC culture  
 XX  
 SQ Sequence 1077 AA;  
 Query Match 100.0%; Score 103; DB 2; Length 1077;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 DB 1033 QRLKYAQEELSNEVLPPPRK 1052  
 RESULT 10  
 AAR32426  
 ID AAR32426 standard; protein; 1091 AA.  
 AC AAR32426;  
 XX 27-AUG-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 17-DEC-2001 (revised)  
 DT 10-JUN-1993 (first entry)  
 XX  
 DE Translated from 5' region of Hepatitis A Virus genomic clone.  
 XX HAV HM-175; chronic liver disease; picornavirus.  
 XX Hepatitis A virus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..711  
 FT /note= "X's correspond to nonsense codons, i.e. this  
 FT region is not an ORF"  
 FT 238..1091  
 FT /label= ORF  
 FT /note= "second putative initiation codon at position 240"  
 XX  
 PN USN7788262-N.  
 XX  
 PD 15-DEC-1992.  
 XX  
 PF 06-NOV-1991; 91US-00788262.  
 XX  
 PR 30-SEP-1983; 88US-00536911.  
 PR 27-SEP-1984; 84US-00654942.  
 PR 06-OCT-1988; 88US-00256135.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Ticehurst JR, Baltimore D, Feinstone SM, Purcell RH;  
 PI Racaniello VR, Baroudy BM, Emerson SU;  
 XX  
 DR WPI; 1993-067429/08.  
 DR N-PSDB; AAQ36934.  
 XX  
 XX Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn. of  
 PT antigen and antibodies.  
 XX  
 PS Disclosure; Fig 7; 65pp; English.  
 XX  
 CC HAV virion RNA was extracted from the livers of marmosets which had been  
 CC inoculated with HAV (the HAV had previously been passaged twice in  
 CC marmosets). The RNA was used to prepare ds cDNA clones by standard  
 CC methods. Clones contg. inserts which hybridised to RNA from HAV-infected  
 CC African Green Monkey kidney cells were selected for further analysis. A  
 CC 7.4kb restriction map (about 99% of the HAV genome) was constructed from  
 CC 5 overlapping inserts. The sequence of the first 3.3kb (approx.) from the  
 CC 5'-terminus was determined. An amino acid sequence was decoded from the

XX 09-OCT-1984; 84EP-00402025.  
 XX  
 PR 14-OCT-1983; 83US-00541836.  
 PR 02-MAR-1984; 84US-00585942.  
 XX  
 PA (MERI ) MERCK & CO INC.  
 XX  
 PI Hughes JV, Scolnick EM, Tomassini JE;  
 XX  
 DR WPI; 1985-100818/17.  
 DR N-PSDB; AAN50274.  
 XX  
 XX New hepatitis A virus surface protein - useful for binding to  
 PT neutralising antibodies to the virus.  
 PS  
 PS Disclosure; Page 17-23; 49pp; English.  
 XX  
 CC VP1 is isolated by solubilisation of the intact virus in an aq. anionic  
 CC surfactant and a reducing agent. The viral proteins are sepd. and the  
 CC protein of molecular wt. 33000 daltons is sepd  
 XX  
 SQ Sequence 993 AA;  
 Query Match 100.0%; Score 103; DB 1; Length 993;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 DB 946 QRLKYAQEELSNEVLPPPRK 965  
 RESULT 9  
 AAW95559  
 ID AAW95559 standard; protein; 1077 AA.  
 AC AAW95559;  
 DT 28-APR-1999 (first entry)  
 XX  
 DE A partial hepatitis A virus (HAV) protein.  
 XX  
 KW Hepatitis A virus protein; HAV; P2 region;  
 KW cell-culture-adapted HAV strain; infection; accelerated growth.  
 XX  
 OS Hepatitis A virus.  
 XX  
 PN US5849562-A.  
 XX  
 PD 15-DEC-1998.  
 XX  
 PF 06-JUN-1995; 95US-00468926.  
 XX  
 PR 30-SEP-1983; 83US-00537911.  
 PR 27-SEP-1984; 84US-00654942.  
 PR 06-OCT-1988; 88US-00256135.  
 PR 06-NOV-1991; 91US-00788262.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Emerson SU, Purcell RH;  
 XX  
 DR WPI; 1999-094412/08.  
 DR N-PSDB; AAX01006.  
 XX  
 XX Chimeric hepatitis A virus strains - with P2 region from cell-culture-  
 PT adapted strain in wild-type genome.  
 XX  
 PS Disclosure; Fig 7A-L; 36pp; English.  
 XX  
 CC The present sequence represents a partial hepatitis A virus (HAV)  
 CC protein. The specification describes a DNA construct consisting of a wild





PR 19-SEP-1984; 84US-00652067.  
 PR 09-SEP-1986; 86US-00905146.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PA Purcell RH, Ticehurst JR, Cohen JL, Emerson SU, Feinstein SM;  
 PI Daemer RJ, Gust ID;  
 XX WPI; 1990-075557/10.  
 DR N-PSDB; RAQ03512.  
 XX Vaccine against hepatitis A virus infection - comprises novel attenuated  
 FT hepatitis A virus strain.  
 XX Claim 1; Fig 1; 18pp; English.  
 XX The attenuated HAV is useful for inducing protective immunity against  
 CC HAV. This strain (pass 35) differs from the wild type HAV HM-175 by  
 CC several nucleotide changes distributed throughout the genome, is  
 CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is  
 CC suitable for use as an HAV vaccine. It is noted that not all the changes  
 CC are necessary for attenuation and use as a vaccine. (Updated on 25-MAR-  
 CC 2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PI field.)  
 CC (Updated on 27-AUG-2003 to correct OS field.) (Updated on 24-OCT-2003 to  
 CC standardise OS field)  
 XX Sequence 2227 AA;  
 SQ  
 Query Match 100.0%; Score 103; DB 2; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 DB 810 QRLKYAQEELSNEVLPPPRK 829  
 RESULT 13  
 AAW34074  
 ID AAW34074 standard; protein; 2227 AA.  
 XX AC AAW34074;  
 XX 17-OCT-2003 (revised)  
 DT 27-APR-1998 (first entry)  
 XX Hepatitis A virus HM-175 protein sequence.  
 XX HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus; infection;  
 KW vaccine.  
 XX Hepatitis A virus; HM-175.  
 OS  
 XX Key Location/Qualifiers  
 FT Protein 1..23  
 FT /label= VP4  
 FT Protein 24..245  
 FT /label= VP2  
 FT Protein 246..491  
 FT /label= VP3  
 FT Protein 492..791  
 FT /label= VP1  
 FT Protein 792..980  
 FT /label= 2A  
 FT Protein 981..1087  
 FT /label= 2B  
 FT Protein 1088..1422  
 FT /label= 2C  
 FT Protein 1423..1496  
 FT /label= 3A  
 FT Protein 1497..1519  
 FT /label= 3B  
 FT Protein 1520..1738

FT Protein /label= 3C  
 FT 1739..2227  
 FT /label= 3D  
 XX WO9740166-A2.  
 XX 30-OCT-1997.  
 XX 18-APR-1997; 97WO-US006506.  
 XX 19-APR-1996; 96US-0015642P.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Raychaudhuri G, Emerson SU, Purcell RH;  
 XX WPI; 1997-535850/49.  
 DR N-PSDB; AAT93023.  
 XX Human attenuated HAV genome containing simian HAV 2C gene - useful as  
 FT vaccines against HAV infection.  
 XX Disclosure; Fig 13A-D; 66pp; English.  
 XX This protein sequence is encoded by the human hepatitis A virus (HAV) HM-  
 CC 175 wild-type genome (see AAT93023). Attenuated strain HAV/7 is obtained  
 CC by passage of HM-175 in African Green Monkey kidney cells. A claimed DNA  
 CC construct (1) comprises a genome of HAV, where the genome is a human  
 CC attenuated HAV genome in which a region of the 2C gene has been replaced  
 CC by a corresponding region from a 2C gene of a simian AGM-27 HAV genome  
 CC (see AAT93024). The region of the 2C gene from AGM-27 contained in the  
 CC construct preferably encodes amino acids 120-328 of the 2C protein, amino  
 CC acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA transcript  
 CC of (1); (2) a cell transfected with (1) or the RNA transcript of (1); (3)  
 CC a HAV genome as above; (4) antibodies to the HAV of (3); and (5) a host  
 CC cell containing the HAV of (3). (1) or its RNA transcript, can be used as  
 CC a vaccine for preventing HAV in a mammal. (1) or the RNA transcript can  
 CC also be used to stimulate the production of protective antibodies in the  
 CC mammal. (Updated on 17-OCT-2003 to standardise OS field)  
 XX SQ Sequence 2227 AA;  
 Query Match 100.0%; Score 103; DB 2; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 DB 810 QRLKYAQEELSNEVLPPPRK 829  
 RESULT 14  
 AAB18609  
 ID AAB18609 standard; protein; 2227 AA.  
 XX AC AAB18609;  
 XX 15-JAN-2001 (first entry)  
 XX Amino acid sequence of live attenuated Hepatitis A virus 4380.  
 DE HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
 KW HAV 4380.  
 XX Hepatitis A virus.  
 OS  
 XX US6113912-A.  
 XX 05-SEP-2000.  
 XX 07-JUN-1995; 95US-00475886.  
 XX 18-SEP-1992; 92US-00947338.  
 PR

PR 17-SEP-1993; 93WO-US008610.  
 PR 17-APR-1995; 95US-00397232.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
 XX  
 DR WPI; 2000-586464/55.  
 DR N-PSDB; AAA75478.  
 XX  
 PT Novel live hepatitis A virus adapted to growth in human fibroblast cell  
 PT line useful as vaccine for protecting humans against hepatitis A virus  
 PT infection, has modified genome compared to wild type.  
 XX  
 XX Disclosure; Col 93-104; 72pp; English.  
 XX  
 PS The present sequence is derived from a live attenuated hepatitis A virus  
 CC (HAV) of the invention, designated HAV 4380. The sequence is produced by  
 CC modifying wild type HAV strain HM-174. The HAV of the invention are  
 CC adapted to growth in the human fibroblast-like cell line MRC-5. The HAV  
 CC is able to propagate in MRC-5 cells and retain appropriate attenuation.  
 CC It is useful as a live vaccine for prophylaxis of hepatitis A in humans  
 CC and other primates  
 XX  
 XX  
 SQ Sequence 2227 AA;  
 Query Match 100.0%; Score 103; DB 3; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 |||||  
 DB 810 QRLKYAQEELSNEVLPPPRK 829  
 |||||  
 RESULT 15  
 AAB18607  
 ID AAB18607 standard; protein; 2227 AA.  
 AC AAB18607;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 DE Amino acid sequence of wild type Hepatitis A virus strain HM-175.  
 XX  
 KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection.  
 XX  
 OS Hepatitis A virus.  
 XX  
 PN US6113912-A.  
 XX  
 PD 05-SEP-2000.  
 XX  
 PF 07-JUN-1995; 95US-00475886.  
 XX  
 PR 18-SEP-1992; 92US-00947338.  
 PR 17-SEP-1993; 93WO-US008610.  
 PR 17-APR-1995; 95US-00397232.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
 XX  
 DR WPI; 2000-586464/55.  
 DR N-PSDB; AAA75478.  
 XX  
 PT Novel live hepatitis A virus adapted to growth in human fibroblast cell  
 PT line useful as vaccine for protecting humans against hepatitis A virus  
 PT infection, has modified genome compared to wild type.  
 XX  
 XX Disclosure; Fig 6A-K; 72pp; English.  
 PS The present sequence is derived from a wild type hepatitis A virus (HAV)

CC strain HM-174. The sequence is modified to produce HAV which are adapted  
 CC to growth in the human fibroblast-like cell line MRC-5. The HAV is able  
 CC to propagate in MRC-5 cells and retain appropriate attenuation. It is  
 CC useful as a live vaccine for prophylaxis of hepatitis A in humans and  
 CC other primates  
 XX  
 XX Sequence 2227 AA;  
 Query Match 100.0%; Score 103; DB 3; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 |||||  
 DB 810 QRLKYAQEELSNEVLPPPRK 829  
 |||||  
 RESULT 16  
 AAB19899  
 ID AAB19899 standard; protein; 2227 AA.  
 XX  
 AC AAB19899;  
 XX  
 DT 18-JUN-2002 (first entry)  
 XX  
 DE Hepatitis A virus (HAV) protein.  
 XX  
 KW Hepatitis A virus; HAV; infection; virucide; fungicide; antibacterial;  
 KW cytostatic; immunostimulant; vaccine; ribavirin; immune response; cancer.  
 XX  
 OS Hepatitis A virus.  
 XX  
 PN WO200213855-A2.  
 XX  
 PD 21-FEB-2002.  
 XX  
 PF 15-AUG-2001; 2001WO-IB001808.  
 XX  
 PR 17-AUG-2000; 2000US-0225767P.  
 PR 29-AUG-2000; 2000US-0229175P.  
 PR 03-NOV-2000; 2000US-00705547.  
 XX  
 PA (TRIP-) TRIPEP AB.  
 XX  
 PI Sallberg M, Hultgren C;  
 XX  
 DR WPI; 2002-241837/29.  
 DR N-PSDB; AAD31766.  
 XX  
 PT Vaccine compositions for treating and preventing disease, preferably  
 PT hepatitis C virus infection, comprises ribavirin and antigen that has  
 PT epitope present in hepatitis C virus.  
 XX  
 PS Claim 11; Page 82-87; 120pp; English.  
 XX  
 CC The invention relates to a composition comprising ribavirin and an  
 CC antigen preferably non structural 3 protein (NS3)/4A fragment of  
 CC hepatitis C virus (HCV) genome or a peptide or nucleic acid of HCV  
 CC sequence. The composition is useful for enhancing an immune response to a  
 CC hepatitis C antigen in humans, domestic, sport or pet species and as  
 CC vaccines for treating and preventing HCV infections. The composition is  
 CC also useful for treating viral, bacterial, fungal diseases and cancer.  
 CC The present sequence is hepatitis A virus (HAV) protein  
 XX  
 XX Sequence 2227 AA;  
 Query Match 100.0%; Score 103; DB 5; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 |||||  
 DB 810 QRLKYAQEELSNEVLPPPRK 829  
 |||||

```

RESULT 17
ABG31729
ID  ABG31729 standard; protein; 2227 AA.
XX  AC
XX  ABG31729;
XX  AC
XX  29-AUG-2003 (revised)
DT  29-NOV-2002 (first entry)
XX  DE
XX  Attenuated Hepatitis A virus (HAV) 4380 polypeptide.
XX  KW
XX  Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast; virucide;
XX  HAV 4380.
XX  OS
XX  Hepatitis A virus; strain HM-175.
XX  PN
XX  US6423318-B1.
XX  PD
XX  23-JUL-2002.
XX  PF
XX  31-AUG-2000; 2000US-00653499.
XX  PR
XX  17-SEP-1993; 93WO-US008610.
XX  PR
XX  17-APR-1995; 95US-00397232.
XX  PR
XX  07-JUN-1995; 95US-00475886.
XX  PA
XX  (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX  PA
XX  (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX  PI
XX  Funkhouser AM, Emerson SU, Purcell RH, D'hondt E;
XX  WPI; 2002-680946/73.
XX  DR
XX  N-PSDB; ABS52789.
XX  DR
XX  US6423318-B1.
XX  PN
XX  23-JUL-2002.
XX  PD
XX  31-AUG-2000; 2000US-00653499.
XX  PF
XX  17-SEP-1993; 93WO-US008610.
XX  PR
XX  17-APR-1995; 95US-00397232.
XX  PR
XX  07-JUN-1995; 95US-00475886.
XX  PA
XX  (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX  PA
XX  (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX  PI
XX  Funkhouser AM, Emerson SU, Purcell RH, D'hondt E;
XX  WPI; 2002-680946/73.
XX  DR
XX  N-PSDB; ABS52789.
XX  DR
XX  New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth
PT  in MRC-5 cells, useful for preparing a vaccine against HAV infection.
XX  PT
XX  Disclosure; Col 93-104; 71pp; English.
XX  PS
XX  The invention relates to a polynucleotide which encodes a hepatitis A
CC  virus (HAV) adapted to growth in MRC-5 cells (human fibroblast-like cell
CC  line). The polynucleotide is useful for preparing a vaccine against
CC  hepatitis A virus infection. This sequence represents an attenuated
CC  hepatitis A virus 4830 polypeptide. (Updated on 29-AUG-2003 to
CC  standardise OS field)
XX  CC
XX  Sequence 2227 AA;
XX  SQ
XX  Query Match 100.0%; Score 103; DB 5; Length 2227;
XX  Best Local Similarity 100.0%; Pred. No. 1.2e-06;
XX  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX  QY
XX  1 QRLKYAQEELSNEVLPPPRK 20
XX  |||||
XX  810 QRLKYAQEELSNEVLPPPRK 829
XX  DB
XX  RESULT 18
ABG31727
ID  ABG31727 standard; protein; 2227 AA.
XX  AC
XX  ABG31727;
XX  AC
XX  29-AUG-2003 (revised)
DT  29-NOV-2002 (first entry)
XX  DE
XX  Wild-type Hepatitis A virus (HAV) strain HM-175 polypeptide.
XX  KW
XX  Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast; virucide.
XX  OS
XX  Hepatitis A virus; strain HM-175.
```

```

XX  US6423318-B1.
XX  PD
XX  23-JUL-2002.
XX  PF
XX  31-AUG-2000; 2000US-00653499.
XX  PR
XX  17-SEP-1993; 93WO-US008610.
XX  PR
XX  17-APR-1995; 95US-00397232.
XX  PR
XX  07-JUN-1995; 95US-00475886.
XX  PA
XX  (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX  PA
XX  (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX  PI
XX  Funkhouser AM, Emerson SU, Purcell RH, D'hondt E;
XX  WPI; 2002-680946/73.
XX  DR
XX  N-PSDB; ABS52787.
XX  DR
XX  New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth
PT  in MRC-5 cells, useful for preparing a vaccine against HAV infection.
XX  PT
XX  Disclosure; Fig 6; 71pp; English.
XX  PS
XX  The invention relates to a polynucleotide which encodes a hepatitis A
CC  virus (HAV) adapted to growth in MRC-5 cells (human fibroblast-like cell
CC  line). The polynucleotide is useful for preparing a vaccine against
CC  hepatitis A virus infection. This sequence represents a hepatitis A virus
CC  strain HM-175 polypeptide. (Updated on 29-AUG-2003 to standardise OS
CC  field)
XX  CC
XX  Sequence 2227 AA;
XX  SQ
XX  Query Match 100.0%; Score 103; DB 5; Length 2227;
XX  Best Local Similarity 100.0%; Pred. No. 1.2e-06;
XX  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX  QY
XX  1 QRLKYAQEELSNEVLPPPRK 20
XX  |||||
XX  810 QRLKYAQEELSNEVLPPPRK 829
XX  DB
XX  RESULT 19
ABU08641
ID  ABU08641 standard; protein; 2227 AA.
XX  AC
XX  ABU08641;
XX  AC
XX  23-OCT-2003 (revised)
DT  03-JUN-2003 (first entry)
XX  DE
XX  Attenuated hepatitis A virus (4380) strain HM-175.
XX  KW
XX  Hepatitis A virus; HAV; virucide; hepatotropic; antiinflammatory;
XX  vaccine; MRC-5 cell; hepatitis infection; ds; attenuated virus.
XX  OS
XX  Hepatitis A virus; strain HM-175.
XX  PN
XX  US2002176869-A1.
XX  PD
XX  28-NOV-2002.
XX  PF
XX  29-APR-2002; 2002US-00135988.
XX  PR
XX  18-SEP-1992; 92US-00947338.
XX  PR
XX  17-SEP-1993; 93WO-US008610.
XX  PR
XX  17-APR-1995; 95US-00397232.
XX  PR
XX  07-JUN-1995; 95US-00475886.
XX  PR
XX  31-AUG-2000; 2000US-00653499.
XX  PA
XX  (FUNK/) FUNKHOUSER A W.
XX  PA
XX  (EMER/) EMERSON S U.
XX  PA
XX  (PURC/) PURCELL R H.
```

PA (DHON/) D'HONDT E.  
XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
XX WPI; 2003-352605/02.  
DR N-PSDB; ABX93475.  
XX New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells,  
PT useful in vaccines for protecting primates against hepatitis infection  
PT and disease.  
XX Disclosure; Page 45-51; 70pp; English.  
XX The invention describes a live hepatitis A virus (HAV) adapted to growth  
CC in MRC-5 cells. The HAV and compositions comprising the HAV are useful as  
CC a vaccine for protecting primates against hepatitis infection and  
CC disease. This is the amino acid sequence of an attenuated human  
CC hepatitis A virus (4380) strain HM-175. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX Sequence 2227 AA;  
SQ Query Match 100.0%; Score 103; DB 6; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 810 QRLKYAQEELSNEVLPPPRK 829  
RESULT 20  
ABU08639  
ID ABU08639 standard; protein; 2227 AA.  
XX AC ABU08639;  
XX 23-OCT-2003 (revised)  
DT 03-JUN-2003 (first entry)  
XX Wild type human hepatitis A virus strain HM-175.  
XX Hepatitis A virus; HAV; virucide; hepatotropic; antiinflammatory;  
XX vaccine; MRC-5 cell; hepatitis infection.  
XX Hepatitis A virus; strain HM-175.  
XX US2002176869-A1.  
XX 28-NOV-2002.  
XX 29-APR-2002; 2002US-00135988.  
XX 18-SEP-1992; 92US-00947338.  
PR 17-SEP-1993; 93WO-US008610.  
PR 17-APR-1995; 95US-00397232.  
PR 07-JUN-1995; 95US-00475886.  
PR 31-AUG-2000; 2000US-00653499.  
XX (FUNK/) FUNKHOUSER A W.  
PA (EMER/) EMERSON S U.  
PA (PURC/) PURCELL R H.  
PA (DHON/) D'HONDT E.  
XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
XX WPI; 2003-352605/02.  
DR N-PSDB; ABX93473.  
XX New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells,  
PT useful in vaccines for protecting primates against hepatitis infection  
PT and disease.  
XX

PS Disclosure; Fig 6; 70pp; English.  
XX The invention describes a live hepatitis A virus (HAV) adapted to growth  
CC in MRC-5 cells. The HAV and compositions comprising the HAV are useful as  
CC a vaccine for protecting primates against hepatitis infection and  
CC disease. This is the amino acid sequence of wild type human hepatitis A  
CC virus strain HM-175. (Updated on 23-OCT-2003 to standardise OS field)  
XX Sequence 2227 AA;  
SQ Query Match 100.0%; Score 103; DB 6; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 810 QRLKYAQEELSNEVLPPPRK 829  
RESULT 21  
ABW00350  
ID ABW00350 standard; protein; 2227 AA.  
XX AC ABW00350;  
XX 15-JAN-2004 (first entry)  
DT Hepatitis A virus protein.  
XX Ribavirin; vaccine; immune response; infection; therapy; immunostimulant;  
XX virucide.  
XX Hepatitis A virus.  
XX US2002136740-A1.  
XX 26-SEP-2002.  
XX 15-AUG-2001; 2001US-00929955.  
PR 17-AUG-2000; 2000US-0225767P.  
PR 29-AUG-2000; 2000US-0229175P.  
XX (SALL/) SALLBERG M.  
PA (HULT/) HULTGREN C.  
XX Sallberg M, Hultgren C;  
XX WPI; 2003-764978/72.  
DR N-PSDB; AAD60867.  
XX Vaccine compositions for treating and preventing disease, preferably  
PT hepatitis C virus infection, comprises ribavirin and antigen that has  
PT epitope present in hepatitis C virus.  
XX Claim 11; Page 45-51; Opp; English.  
XX The invention relates to a composition comprising ribavirin and an  
CC antigen, where the antigen is derived from a hepatitis virus. The vaccine  
CC is useful in enhancing the immune response to a hepatitis C antigen where  
CC the composition is delivered to an animal identified as requiring an  
CC enhanced immune response. The vaccine is useful in the treatment and  
CC prevention of hepatitis C infection. The present sequence is Hepatitis A  
CC virus protein  
XX Sequence 2227 AA;  
SQ Query Match 100.0%; Score 103; DB 7; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 810 QRLKYAQEELSNEVLPPPRK 829

Db 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 22

AAE18608

ID AAE18608 standard; protein; 2227 AA.

XX AC AAE18608;

XX DT 15-JAN-2001 (first entry)

XX DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.

XX KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;

XX KW P-35 virus.

XX OS Hepatitis A virus.

XX PN US6113912-A.

XX PD 05-SEP-2000.

XX PF 07-JUN-1995; 95US-00475886.

XX PR 18-SEP-1992; 92US-00947338.

XX PR 17-SEP-1993; 93WO-US008610.

XX PR 17-APR-1995; 95US-00397232.

XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX PI D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;

XX XX WPI; 2000-586464/55.

XX DR N-PSDB; AAA75477.

XX PT Novel live hepatitis A virus adapted to growth in human fibroblast cell

XX PT line useful as vaccine for protecting humans against hepatitis A virus

XX PT infection, has modified genome compared to wild type.

XX PS Disclosure; Col 67-78; 72pp; English.

XX CC The present sequence is derived from passage 35 of a wild type hepatitis

XX CC A virus (HAV) strain HM-174. The resulting virus is designated P-35

XX CC virus. The sequence is modified to produce HAV which are adapted to

XX CC growth in the human fibroblast-like cell line MRC-5. The HAV is able to

XX CC propagate in MRC-5 cells and retain appropriate attenuation. It is useful

XX CC as a live vaccine for prophylaxis of hepatitis A in humans and other

XX CC primates

XX SQ Sequence 2227 AA;

Query Match 95.1%; Score 98; DB 3; Length 2227;

Best Local Similarity 95.0%; Pred. No. 7.2e-06;

Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20

|||||||:|||||||

Db 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 23

ABG31728

ID ABG31728 standard; protein; 2227 AA.

XX AC ABG31728;

XX DT 29-NOV-2002 (first entry)

XX DE Hepatitis A virus mutant strain HM-175/7 (pHAV/7) polypeptide.

XX KW Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast;

XX KW virucide; mutant; pHAV/7; mutcin.

OS Hepatitis A virus; strain HM-175.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 764 /note= "Wild-type Glu substituted by Val"

FT Misc-difference 821 /note= "Wild-type Asn substituted by Ser"

FT Misc-difference 963 /label= Wild-type Lys substituted by Arg

FT Misc-difference 1052 /note= "Wild-type Ala substituted by Val"

FT Misc-difference 1062 /note= "Wild-type Gly substituted by Ala"

FT Misc-difference 1118 /note= "Wild-type Lys substituted by Met"

FT Misc-difference 1151 /note= "Wild-type Glu substituted by Lys"

FT Misc-difference 1163 /note= "Wild-type Phe substituted by Ser"

FT Misc-difference 1277 /note= "Wild-type Val substituted by Ile"

FT Misc-difference 1500 /note= "Wild-type His substituted by Tyr"

FT Misc-difference 1805 /note= "Wild-type Asp substituted by Asn"

FT Misc-difference 1930 /note= "Wild-type Ser substituted by Thr"

XX US6423318-B1.

XX 23-JUL-2002.

XX 31-AUG-2000; 2000US-00653499.

XX 17-SEP-1993; 93WO-US008610.

XX 17-APR-1995; 95US-00397232.

XX 07-JUN-1995; 95US-00475886.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;

XX WPI; 2002-680946/73.

XX N-PSDB; ABS52788.

XX New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth

XX in MRC-5 cells, useful for preparing a vaccine against HAV infection.

XX Example 3; Col 67-78; 71pp; English.

XX The invention relates to a polynucleotide which encodes a hepatitis A

XX virus (HAV) adapted to growth in MRC-5 cells (human fibroblast-like cell

XX line). The polynucleotide is useful for preparing a vaccine against

XX hepatitis A virus infection. This sequence represents a hepatitis A virus

XX mutant strain HM-175/7 (pHAV/7) polypeptide

XX Sequence 2227 AA;

Query Match 95.1%; Score 98; DB 5; Length 2227;

Best Local Similarity 95.0%; Pred. No. 7.2e-06;

Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20

|||||||:|||||||

Db 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 24

ABU08640

ID ABU08640 standard; protein; 2227 AA.

XX

AC ABU08640;

XX 23-OCT-2003 (revised)

DT 03-JUN-2003 (first entry)

XX Attenuated (pass35) hepatitis A virus strain HM-175.

XX Hepatitis A virus; HAV; viruicide; hepatotropic; antiinflammatory;

KW vaccine; MRC-5 cell; hepatitis infection; ds; attenuated virus.

XX Hepatitis A virus; strain HM-175.

OS US2002176869-A1.

XX 28-NOV-2002.

XX 29-APR-2002; 2002US-00135988.

XX 18-SEP-1992; 92US-00947338.

PR 17-SEP-1993; 93WO-US008610.

PR 17-APR-1995; 95US-00397232.

PR 07-JUN-1995; 95US-00475886.

PR 31-AUG-2000; 2000US-00653499.

XX (FUNK/) FUNKHOUSER A W.

PA (EMER/) EMERSON S U.

PA (PURC/) PURCELL R H.

PA (DHON/) D'HONDT E.

XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;

PI WPI; 2003-352605/02.

XX N-PSDB; ABX93474.

DR New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells,

XX useful in vaccines for protecting primates against hepatitis infection

PT and disease.

XX Example 3; Fig 6; 70pp; English.

PS The invention describes a live hepatitis A virus (HAV) adapted to growth

CC in MRC-5 cells. The HAV and compositions comprising the HAV are useful as

CC a vaccine for protecting primates against hepatitis infection and

CC disease. This is the amino acid sequence of an attenuated (pass 35)

CC human hepatitis A virus strain HM-175. (Updated on 23-OCT-2003 to

CC standardise OS field)

XX Sequence 2227 AA;

SQ Query Match 95.1%; Score 98; DB 6; Length 2227;

Best Local Similarity 95.0%; Pred. No. 7.2e-06;

Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20

DB 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 25

AA15629

ID AA15629 standard; protein; 839 AA.

XX AA15629;

XX 24-OCT-2003 (revised)

DT 27-AUG-2003 (revised)

DT 25-MAR-2003 (revised)

DT 17-DEC-2001 (revised)

DT 17-MAR-1992 (first entry)

XX Capsid region of cyno-HAV isolate CY-145.

DE Hepatitis A virus; cynomolgus; HAV; monkey; vaccine; macaque.

XX

XX Viruses.

OS Key Location/Qualifiers

XX 1..245

FT /label= VP0

FT Cleavage-site 245..246

FT Protein 246..491

FT /label= VP3

FT Active-site 315

FT Cleavage-site 491..492

FT Protein 492..791

FT /label= VP1

FT Active-site 593

FT Cleavage-site 791..792

FT Protein 792

FT /label= P2

FT /note= "incomplete"

XX USN7678828-N.

PN 12-NOV-1991.

PD 03-APR-1991; 91US-00678828.

PF 03-APR-1991; 91US-00678828.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.

PR Hainan OV, Margolis HS, Robertson BH, Brinton MH, Ebert JW;

XX WPI; 1991-376737/51.

DR N-PSDB; AAQ15180.

XX Hepatitis A virus isolates and DNA - used to prepare vaccines for

PT preventing hepatitis A virus infection.

XX Disclosure; Fig 3; 23pp; English.

PS The sequence was deduced from the nucleotide sequence obt'd. by PCR

CC amplification of cyno-HAV viral RNA obt'd. from the stool of a cynomolgus

CC monkey with serologically and histologically confirmed spontaneous

CC hepatitis A. The sequence differs from the human HAV isolate HMI75

CC (Cohen, J.I., et al. (1987) Proc. Natl. Acad. Sci. USA 84, 2497-2501),

CC mainly in the VP3 and VP1 proteins. The Gln-Val pair at the VP3-VP1

CC cleavage site in the human isolate is replaced by a Gln-Thr pair in the

CC cyno-HAV. The other two cleavage sites are the same. Two residues have

CC been identified as part of the immuno- dominant region (see feature

CC table) and are different to those in the same position in human HAV. The

CC protein and peptides derived from it can be used in the prep'n. of

CC vaccines for the prevention of HAV infection. See also AAR15056. (Note:

CC Government-owned NTIS applications to correct the patent number format of US

CC granted patent numbers. For further information please visit the Derwent

CC web site at [www.derwent.com/dwpi/updates/ntis.us.html](http://www.derwent.com/dwpi/updates/ntis.us.html).) (Updated on 25-

CC MAR-2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS

CC field.) (Updated on 24-OCT-2003 to standardise OS field)

XX Sequence 839 AA;

SQ Query Match 91.3%; Score 94; DB 2; Length 839;

Best Local Similarity 85.0%; Pred. No. 1.1e-05;

Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20

DB 809 QRLKYAQEELSNEVLPPPRK 828

RESULT 26

ABB68507

ID ABB68507 standard; protein; 150 AA.

XX

```

AB 68507;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 32313.
DE Drosophila; developmental biology; cell signalling; insecticide;
XX Drosophila; developmental biology; cell signalling; insecticide;
XX Drosophila melanogaster.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-556860/75.
XX N-PSDB; ABL12610.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX Disclosure; SEQ ID NO 32313; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX sequences (AB101840-AB16175) and the encoded proteins (AB157737-
XX AB172072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 150 AA;
XX
Query Match 48.5%; Score 50; DB 4; Length 150;
Best Local Similarity 52.4%; Pred. NO. 10;
Matches 11; Conservative 3; Mismatches 3; Indels 4; Gaps 1;
QY 1 ORLKXQ----BELSNEVLPP 17
DB 32 RELKYRQRRDVGSIANEYLPP 52
XX
RESULT 27
AAW20624
ID AAW20624 standard; protein; 444 AA.
XX
AC AAW20624;
XX
DT 14-JUL-1997 (first entry)
XX
DE H. pylori cytoplasmic protein, 02aell612orf25.
XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX identification; binding compound; bacterium; life cycle; activator;
XX bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
XX diagnosis.
XX Helicobacter pylori.
XX OS
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WO9640893-A1.

```

---

```

XX 19-DEC-1996.
XX 06-JUN-1996; 96WO-US009122.
XX 07-JUN-1995; 95US-00487032.
XX 01-APR-1996; 96US-00630405.
XX (ASTR ) ASTRA AB.
XX Smith D, Berglinth OT, Mellgaard BL;
XX WPI; 1997-052306/05.
XX N-PSDB; AAT67877.
XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
XX useful for vaccines to treat or prevent H. pylori infection, and to
XX detect Helicobacter.
XX Claim 61; Page 1050-51; 1481pp; English.
XX This sequence represents a H. pylori cytoplasmic protein. The protein may
XX be used in a vaccine to prevent or treat H. pylori infection or to
XX identify H. pylori polypeptide binding compounds, useful as potential H.
XX pylori life cycle activators or inhibitors. The genomic sequence of H.
XX pylori (ATCC 55679) was determined from overlapping contigs generated by
XX mechanically shearing the bacterial DNA. The sequences were analysed for
XX ORF of at least 180 nucleotides, and the predicted coding regions defined
XX by computer evaluation. To identify likely H. pylori antigens for vaccine
XX development, the amino acid sequences predicted from various ORF were
XX analysed for significant homology to other known or exported membrane
XX proteins. Having identified and determined the sequences of interest,
XX particular regions can be isolated from H. pylori by PCR amplification
XX for recombinant polypeptide production, e.g. in E. coli hosts
XX Sequence 444 AA;
XX
Query Match 46.8%; Score 48; DB 2; Length 444;
Best Local Similarity 75.0%; Pred. NO. 64;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 9 ELSNEVLPPPRK 20
DB 416 EADNEELPPPRK 427
XX
RESULT 28
ABBS9692
ID ABBS9692 standard; protein; 1176 AA.
XX
AC ABBS9692;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 5868.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX
DT 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX

```

XX WPI; 2001-656860/75.  
DR N-PSDB; ABL03795.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 5868; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1176 AA;  
Query Match 46.6%; Score 48; DB 4; Length 1176;  
Best Local Similarity 56.2%; Pred. No. 1.8e+02;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPP 16  
:|||||:|:|  
DB 335 KRKYAKELQKEMLP 350  
RESULT 29  
ABB68504  
ID ABB68504 standard; protein; 221 AA.  
XX  
AC ABB68504;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 32304.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL12607.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 32304; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 221 AA;  
Query Match 45.6%; Score 47; DB 4; Length 221;  
Best Local Similarity 47.6%; Pred. No. 44;  
Matches 10; Conservative 4; Mismatches 3; Indels 4; Gaps 1;  
QY 1 QRLKYAQEELSNEVLPP 17  
:|||||:|:|  
DB 128 RLKRYQRDVSDIANEVLPP 148  
RESULT 30  
AAW42923  
ID AAW42923 standard; peptide; 20 AA.  
XX  
AC AAW42923;  
XX  
DT 28-APR-1998 (first entry)  
XX  
DE Immunogenic Hepatitis A virus peptide YK-1316.  
XX  
KW Immunogenic peptide; immunogenic epitope; P2A protein; immune response;  
KW antibody.  
XX  
OS Synthetic.  
OS Hepatitis A virus.  
XX  
PN WO9740147-A1.  
XX  
PD 30-OCT-1997.  
XX  
PF 18-APR-1997; 97WO-US006891.  
XX  
PR 19-APR-1996; 96US-0015644P.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI; 1997-535831/49.  
XX  
PT Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
PT response to HAV in a mammal or to detect the presence of antibodies  
PT against HAV in a mammal.  
XX  
PS Claim 18; Page 112; 140pp; English.  
XX  
CC Peptides AAW42922-30 are immunogenic peptides corresponding to  
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
CC substantially similar to a portion of the amino acid sequence of the P2A  
CC protein of HAV corresponding to amino acids 792-980. The present peptide  
CC is derived from amino acids 799-818, and has a reactivity of 41.7% with  
CC acute sera. Compositions containing the peptides can be used to induce an  
CC immune response to HAV in a mammal. The peptides can also be used to  
CC detect the presence of antibodies against HAV in mammalian serum. The  
CC peptides can also be used to make an antibody against HAV by  
CC administering the peptide to a mammal  
XX  
SQ Sequence 20 AA;  
Query Match 43.7%; Score 45; DB 2;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEE 9  
:|||||:



Db 12 QRLKYAEE 20

Search completed: May 11, 2004, 13:37:28  
Job time : 50 secs

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:30:01 ; Search time 10.5 Seconds

183.222 Million cell updates/sec  
(without alignments)

Title: US-09-171-432A-41

Perfect score: 103

Sequence: 1 QRLKYAQEELSNEVLPPRK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Minimum	DB seq	length:	0
Maximum	DB seq	length:	2000000000

Post-processing: Minimum Match 0%

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 100 summaries

Database : PIR 78:\*

```
1:  pir1:
```

```
2: pir2:*
```

3: p1r3:

2  
4  
4  
4  
2

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	103	100.0	341	2	S04137	genome polyprotein
2	103	100.0	852	1	GNNYHA	genome polyprotein
3	103	100.0	1358	2	A03905	genome polyprotein
4	103	100.0	2227	1	GNNYHM	genome polyprotein
5	103	100.0	2227	1	GNNYHR	genome polyprotein
6	103	100.0	2227	1	GNNYHB	genome polyprotein
7	98	95.1	2327	1	GNNYMK	genome polyprotein
8	97	94.2	2230	1	GNNYSA	genome polyprotein
9	94	91.3	839	1	GNNYS2	genome polyprotein
10	56	54.4	346	2	S74448	regulatory protein
11	52	50.5	859	2	T43701	DNA-directed RNA p
12	52	50.5	1193	2	E88445	protein C26E5.4 [i
13	51	49.5	1119	2	T50995	related to cytoske
14	49.5	48.1	443	2	E82046	proteinase HalVU,
15	49	47.6	55	2	PQ0433	genome polyprotein
16	49	47.6	56	2	PQ0434	genome polyprotein
17	49	47.6	56	2	PQ0428	genome polyprotein
18	49	47.6	56	2	PQ0427	genome polyprotein
19	49	47.6	56	2	PQ0429	genome polyprotein
20	49	47.6	56	2	PQ0432	genome polyprotein
21	49	47.6	56	2	PQ0430	genome polyprotein
22	49	47.6	1174	2	S28976	DNA-directed RNA p
23	48	46.6	442	2	A71969	probable histidine
24	48	46.6	1176	2	A27826	DNA-directed RNA p
25	48	46.6	1191	2	S65068	DNA-directed RNA p
26	47	45.6	592	1	LLBY	actin-binding prot
27	47	45.6	6642	2	T29757	protein UNC-89 - C
28	46	44.7	338	2	I56893	transcription beta
29	45	44.7	480	1	FWPU1B	I1S globulin beta

## ALIGNMENTS

RESULT 1  
S04137  
genome polyprotein - human hepatitis A virus (strain LCDG-1) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
C:Accession: S04137  
R:Andonov, A.P.; Lau, P.C.K.; Chaudhary, R.  
Nucleic Acids Res. 17, 3594, 1989  
A:Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A virus  
A:Reference number: S04137; MUID:89263805; PMID:2542903  
A:Accession: S04137  
A:Molecule type: mRNA  
A:Residues: 1-341 <AND>  
A:Cross-references: EMBL:X14666; NID:g62301; PIDN:CAA32794.1; PID:g4377576  
C:Genetics:  
A:Gene: VP1  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; polyprotein  
F:2-340/Product: coat protein 1D (VP1) #status predicted <MAT>  
  
Query Match 100.0%; Score 103; DB 2; Length 341;  
Best Local Similarity 100.0%; Pred. No. 1.8e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ORLKYAQEELSNEVLPPPRK 20  
|||||  
DB 314 ORLKYAQEELSNEVLPPPRK 333  
  
RESULT 2  
GNVYHA  
genome polyprotein - human hepatitis A virus (strain CR326) (fragment)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03904  
R:Linemeyer, D.L.; Menke, J.G.; Martin-Gallardo, A.; Hughes, J.V.; Young, A.; Mitra, S.K.  
J. Virol. 54, 247-255, 1985  
A:Title: Molecular cloning and partial sequencing of hepatitis A viral cDNA.  
A:Reference number: A03904; MUID:85185648; PMID:2985793  
A:Accession: A03904  
A:Molecule type: genomic RNA  
A:Residues: 1-852 <LIN>  
A:Cross-references: EMBL:M10033; NID:g329592; PIDN:AAA45470.1; PID:g329593  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-852/Product: coat protein 2A (fragment) #status predicted <C2A>  
  
Query Match 100.0%; Score 103; DB 1; Length 852;  
Best Local Similarity 100.0%; Pred. No. 5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ORLKYAQEELSNEVLPPPRK 20  
|||||  
DB 810 ORLKYAQEELSNEVLPPPRK 829  
  
RESULT 3  
A03905  
genome polyprotein (version 2) - human hepatitis A virus (fragments)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; protein  
C:Species: human hepatitis A virus  
C:Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 15-Nov-1996  
C:Accession: A03905  
R:Baroudy, B.M.; Ticehurst, J.R.; Miele, T.A.; Maizel Jr., J.V.; Purcell, R.H.; Feinstone  
Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985

A:Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA  
A:Reference number: A03905; MUID:85166289; PMID:2984684  
A:Accession: A03905  
A:Molecule type: genomic RNA  
A:Residues: 1-1358 <BAR>  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-852/Product: coat protein 2A (fragment) #status predicted <C2A>  
F:855-1358/Product: protein 3D (RNA polymerase) (fragment) #status predicted <C3D>  
  
Query Match 100.0%; Score 103; DB 2; Length 1358;  
Best Local Similarity 100.0%; Pred. No. 8.5e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ORLKYAQEELSNEVLPPPRK 20  
|||||  
DB 810 ORLKYAQEELSNEVLPPPRK 829  
  
RESULT 4  
GNVYHM  
genome polyprotein - human hepatitis A virus (strain HM-175, wild type)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro  
B:RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999  
C:Accession: A25981  
R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.  
J. Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with di  
A:Reference number: A25981; MUID:87061253; PMID:3023706  
A:Accession: A25981  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <CH>  
A:Cross-references: EMBL:M14707; NID:g329582; PIDN:AAA45465.1; PID:g329583  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-491/Product: coat protein 1C #status predicted <VP3>  
F:492-791/Product: coat protein 1D #status predicted <VP1>  
F:792-980/Product: core protein 2A #status predicted <C2A>  
F:981-1087/Product: core protein 2B #status predicted <C2B>  
F:1088-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1496/Product: protein 3A #status predicted <C3A>  
F:1497-1519/Product: protein 3B #status predicted <C3B>  
F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>  
  
Query Match 100.0%; Score 103; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;  
  
QY 1 ORLKYAQEELSNEVLPPPRK 20  
|||||  
DB 810 ORLKYAQEELSNEVLPPPRK 829  
  
RESULT 5  
GNVYHR  
genome polyprotein - human hepatitis A virus  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core pro  
NA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A03903  
R:Najarian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van Nest  
Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985

A;Title: Primary structure and gene organization of human hepatitis A virus.  
A;Reference number: A03903; MUID:85190549; PMID:2986127  
A;Accession: A03903  
A;Molecule type: genomic RNA  
A;Residues: 1-2227 <RNA>  
A;Cross-references: GB:K02990; NID:G329596; PIDN:AAA45472.1; PID:G329597  
C;Superfamily: hepatitis A virus genome polypeptide  
C;Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltransferase  
F;1-245/Product: coat protein 1A #status predicted <C1A>  
F;245-491/Product: coat protein 1B #status predicted <C1B>  
F;492-836/Product: coat protein 1C #status predicted <C1C>  
F;837-980/Product: coat protein 2A #status predicted <C2A>  
F;981-1076/Product: core protein 2B #status predicted <C2B>  
F;1077-1422/Product: core protein 2C #status predicted <C2C>  
F;1423-1484/Product: protein 3A #status predicted <C3A>  
F;1485-1507/Product: protein 3B #status predicted <C3B>  
F;1508-1678/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F;1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>  
Query Match 100.0%; Score 103; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 QRLKYAQEELSNEVLPPPRK 20  
Db 810 QRLKYAQEELSNEVLPPPRK 829  
RESULT 6  
GNNYH  
N;Genome polypeptide - human hepatitis A virus (strain MB8)  
N;Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core protein 2A; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C;Species: human hepatitis A virus  
A;Note: host Homo sapiens (man)  
C;Accession: J50303  
C;Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 10-May-1996  
R;Paul, A.V.; Tada, H.; von der Helm, K.; Wessel, T.; Kiehn, R.; Wimmer, E.; Deinhardt, Virus Res. 8, 153-171, 1987  
A;Title: The entire nucleotide sequence of the genome of human hepatitis A virus (isolat  
A;Reference number: J50303; MUID:84045071; PMID:2823500  
A;Accession: J50303  
A;Molecule type: genomic RNA  
A;Residues: 1-2227 <PAU>  
A;Cross-references: EMBL:M20273  
C;Superfamily: hepatitis A virus genome polypeptide  
C;Keywords: coat protein; core protein; cysteine proteinase; genome-linked protein; hydrolase  
F;1-23/Product: coat protein 1A #status predicted <VP4>  
F;24-246/Product: coat protein 1B #status predicted <VP2>  
F;247-491/Product: coat protein 1C #status predicted <VP3>  
F;492-836/Product: coat protein 1D #status predicted <VP1>  
F;837-980/Product: core protein 2A #status predicted <P2A>  
F;981-1108/Product: core protein 2B #status predicted <P2B>  
F;1109-1438/Product: core protein 2C #status predicted <P2C>  
F;1439-1496/Product: protein 3A #status predicted <P3A>  
F;1497-1519/Product: genome-linked protein VPg #status predicted <VPG>  
F;1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F;1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>  
Query Match 100.0%; Score 103; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 QRLKYAQEELSNEVLPPPRK 20  
Db 810 QRLKYAQEELSNEVLPPPRK 829  
RESULT 7  
GNNYK  
N;Genome polypeptide - human hepatitis A virus (strain HM-175/7MK-5, attenuated HAV)  
N;Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core protein 2B; RNA polymerase (EC 2.7.7.48), protein 3D

C;Species: human hepatitis A virus  
A;Note: host Homo sapiens (man)  
C;Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 16-Jul-1999  
C;Accession: A94149; A25914; A94508  
R;Cohen, J.I.; Rosenblum, B.; Ticehurst, J.R.; Daemer, R.J.; Feinstein, S.M.; Purcell, R. Proc. Natl. Acad. Sci. U.S.A. 84, 2497-2501, 1987  
A;Title: Complete nucleotide sequence of an attenuated hepatitis A virus: comparison with  
A;Reference number: A94149; MUID:87175701; PMID:3031686  
A;Accession: A94149  
A;Status: nucleic acid sequence not shown  
A;Molecule type: genomic RNA  
A;Residues: 1-2227 <COH>  
A;Cross-references: EMBL:M16632; NID:G329594; PIDN:AAA45471.1; PID:G329595  
A;Note: submitted to GenBank, August 1987  
C;Superfamily: hepatitis A virus genome polypeptide  
C;Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltransferase  
F;1-245/Product: coat protein 1A #status predicted <P1A>  
F;246-491/Product: coat protein 1B #status predicted <P1B>  
F;492-836/Product: coat protein 1C #status predicted <P1C>  
F;837-980/Product: core protein 2A #status predicted <P2A>  
F;981-1076/Product: core protein 2B #status predicted <P2B>  
F;1077-1422/Product: core protein 2C #status predicted <P2C>  
F;1423-1484/Product: protein 3A #status predicted <P3A>  
F;1485-1507/Product: protein 3B #status predicted <P3B>  
F;1508-1678/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F;1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>  
Query Match 95.1%; Score 98; DB 1; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 9.2e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 QRLKYAQEELSNEVLPPPRK 20  
Db 810 QRLKYAQEELSNEVLPPPRK 829  
RESULT 8  
GNNYSA  
N;Genome polypeptide - simian hepatitis A virus (strain AGM-27)  
N;Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core protein 2A; RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C;Species: simian hepatitis A virus  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jun-2000  
C;Accession: A30470; S04885; S03965  
R;Tsarev, S.A.  
A;Note: submitted to JIPID, April 1991  
A;Reference number: A30470  
A;Accession: A30470  
A;Molecule type: genomic RNA  
A;Residues: 1-2230 <TSA>  
A;Cross-references: GB:D00924; NID:G222597; PIDN:BAA00766.1; PID:G222598  
R;Tsarev, S.A.; Emerson, S.U.; Balayan, M.S.; Ticehurst, J.; Purcell, R.H. J. Gen. Virol. 72, 1677-1683, 1991  
A;Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure and  
A;Reference number: JQ1080; MUID:91311420; PMID:1649901  
A;Contents: annotation  
A;Note: neither amino acid nor nucleotide sequence is given  
R;Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhikov  
submitted to the EMBL Data Library, May 1989  
A;Reference number: S04885  
A;Accession: S04885  
A;Molecule type: genomic RNA  
A;Residues: 1750-2164 <BAL1>  
A;Cross-references: EMBL:X15461; NID:G61971; PIDN:CRA33490.1; PID:G930268  
R;Balayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsarev, S.A.; Sverdlov, E.D.; Chizhikov  
FEBS Lett. 247, 425-428, 1989  
A;Title: Variations in genome fragments coding for RNA polymerase in human and simian he  
A;Reference number: S03965; MUID:89232168; PMID:2541023  
A;Accession: S03965  
A;Molecule type: genomic RNA  
A;Residues: 1960-2164 <BAL2>  
A;Cross-references: EMBL:X15461  
C;Superfamily: hepatitis A virus genome polypeptide  
C;Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltransferase

F;1-27/Product: coat protein 1A #status predicted <C1A>  
F;28-249/Product: coat protein 1B #status predicted <C1B>  
F;250-495/Product: coat protein 1C #status predicted <C1C>  
F;496-795/Product: coat protein 1D #status predicted <C1D>  
F;796-984/Product: coat protein 2A #status predicted <C2A>  
F;985-1091/Product: coat protein 2B #status predicted <C2B>  
F;1092-1426/Product: coat protein 2C #status predicted <C2C>  
F;1427-1498/Product: protein 3A #status predicted <P3A>  
F;1499-1521/Product: protein 3B #status predicted <P3B>  
F;1522-1741/Product: protein 3C #status predicted <P3C>  
F;1742-2230/Product: protein 3D #status predicted <P3D>

Query Match 94.2%; Score 97; DB 1; Length 2230;  
Best Local Similarity 90.0%; Pred. No. 1.3e-06;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 814 QRLKYAMEELSNEILPPPRK 833

RESULT 9  
GNYS2  
genome polyprotein - simian hepatitis A virus (strain CY-145) (fragment)  
N;Contains: amino end of core protein 2A; coat protein 1A; coat protein 1B; coat protein 1C;Species: simian hepatitis A virus  
A;Note: host Macaca fascicularis (cynomolgus macaque)  
C;Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 16-Jul-1999  
C;Accession: JQ1180  
R;Nainan, O.V.; Margolis, H.S.; Robertson, B.H.; Balayan, M.; Brinton, M.A.  
J. Gen. Virol. 72, 1685-1689, 1991  
A;Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus macaque  
A;Reference number: JQ1180; MUID:91311421; PMID:1649902  
A;Accession: JQ1180  
A:Molecule type: Genomic RNA  
A;Residues: 1-839 <NAI>  
A;Cross-references: NID:G329599; PIDN:AAA45473.1; PID:9555083  
C;Superfamily: hepatitis A virus genome polyprotein  
C;Keywords: coat protein; core protein; glycoprotein; polyprotein  
F;1-23/Product: coat protein 1A #status predicted <VP0>  
F;24-245/Product: coat protein 1B #status predicted <VP3>  
F;246-491/Product: coat protein 1C #status predicted <VP1>  
F;492-839/Product: coat protein 2A (fragment) #status predicted <P2P>  
F;261,312,728,756/Binding site: carbohydrate (Asn) #status predicted

Query Match 91.3%; Score 94; DB 1; Length 839;  
Best Local Similarity 85.0%; Pred. No. 1.3e-06;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 809 QRLKYAMEELSNEILPPPRK 828

RESULT 10  
S74448  
regulatory protein pcr - Synecocystis sp. (strain PCC 6803)  
N;Alternate names: protein sll1408  
C;Species: Synecocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 18-Aug-2000  
C;Accession: S74448  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Oikawa, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, C. J. Mol. Biol. 272, 109-136, 1996  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S74448  
A;Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
R;Residues: 1-346 <KAN>

RESULT 13  
T50995  
related to cytoskeleton assembly control protein SLA1 [imported] - Neurospora crassa  
N:Alternate names: protein B7F18.140  
C:Species: Neurospora crassa  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, July 2000  
A:Reference number: Z5286  
A:Accession: T50995  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1119 <SCH>  
A:Cross-references: ENBL:AL389891; GSPDB:GN00116; NCSP:B7F18.140  
A:Experimental source: BAC clone B7F18; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B7F18.140  
A:Map position: 6  
A:Introns: 66/3; 123/2; 495/1

Query Match 49.5%; Score 51; DB 2; Length 1119;  
Best Local Similarity 55.6%; Pred. No. 11;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 QRLKYAQEELNEVLPPP 18  
DB 165 QRSVASEDYENVRSP 182

RESULT 14  
B82046  
proteinase HslVU, ATPase subunit HslU VC2674 [imported] - Vibrio cholerae (strain N16961)  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: B82046  
R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.B.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: B82035; MUID:20406833; PMID:10952301  
A:Accession: B82046  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-443 <HEI>  
A:Cross-references: GB:AE004333; GB:AE003852; NID:95657266; PIDN:AAF95815.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC2674  
A:Map position: 1  
C:Superfamily: heat shock protein hslU; FtsH/SEC18/CDC48-type ATP-binding domain homolog

Query Match 48.1%; Score 49.5; DB 2; Length 443;  
Best Local Similarity 41.7%; Pred. No. 6.8;  
Matches 10; Conservative 6; Mismatches 3; Indels 5; Gaps 1;

QY 1 QRLKYAQEELSNE-----VLPPPR 19  
DB 117 EKVFRAEELAEERVLALLPPPR 140

RESULT 15  
PQ0433  
genome polyprotein - human hepatitis A virus (strain PA21) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0433  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel, J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0433

A:Molecule type: mRNA  
A:Residues: 1-55 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 47.6%; Score 49; DB 2; Length 55;  
Best Local Similarity 100.0%; Pred. No. 0.77;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEEL 10  
DB 46 QRLKYAQEEL 55

RESULT 16  
PQ0434  
genome polyprotein - human hepatitis A virus (strain KPH) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0434  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel, J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0434  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 47.6%; Score 49; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEEL 10  
DB 47 QRLKYAQEEL 56

RESULT 17  
PQ0428  
genome polyprotein - human hepatitis A virus (strain TKM002) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0428  
R:Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel, J. Gen. Virol. 73, 1365-1377, 1992  
A:Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A:Reference number: PQ0427; MUID:92300330; PMID:1318940  
A:Accession: PQ0428  
A:Molecule type: mRNA  
A:Residues: 1-56 <ROB>  
A:Note: this protein is from the VP1/2A Junction region  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein

Query Match 47.6%; Score 49; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEEL 10  
DB 47 QRLKYAQEEL 56

RESULT 18  
PQ0427  
genome polyprotein - human hepatitis A virus (strain EP-35.730) (fragment)  
C:Species: human hepatitis A virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PQ0427

R;Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A;Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A;Reference number: PQ0427; MUID:92300330; PMID:1318940  
A;Accession: PQ0427  
A;Molecule type: mRNA  
A;Residues: 1-56 <ROB>  
C;Comment: This protein is from the VP1/2A Junction region.  
C;Superfamily: hepatitis A virus genome polyprotein  
C;Keywords: coat protein; core protein; polyprotein

Query Match 47.6%; Score 49; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEEL 10  
| | | | | | | |  
Db 47 QRLKYAQEEL 56

RESULT 19  
PQ0429  
genome polyprotein - human hepatitis A virus (strain PRC16) (fragment)  
C;Species: human hepatitis A virus  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C;Accession: PQ0429  
R;Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A;Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A;Reference number: PQ0427; MUID:92300330; PMID:1318940  
A;Accession: PQ0429  
A;Molecule type: mRNA  
A;Residues: 1-56 <ROB>  
A;Note: this protein is from the VP1/2A Junction region  
C;Superfamily: hepatitis A virus genome polyprotein  
C;Keywords: coat protein; core protein; polyprotein

Query Match 47.6%; Score 49; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEEL 10  
| | | | | | | |  
Db 47 QRLKYAQEEL 56

RESULT 20  
PQ0432  
genome polyprotein - human hepatitis A virus (strain CF-53) (fragment)  
C;Species: human hepatitis A virus  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C;Accession: PQ0432  
R;Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A;Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A;Reference number: PQ0427; MUID:92300330; PMID:1318940  
A;Accession: PQ0432  
A;Molecule type: mRNA  
A;Residues: 1-56 <ROB>  
A;Note: this protein is from the VP1/2A Junction region  
C;Superfamily: hepatitis A virus genome polyprotein  
C;Keywords: coat protein; core protein; polyprotein

Query Match 47.6%; Score 49; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEEL 10  
| | | | | | | |  
Db 47 QRLKYAQEEL 56

RESULT 21  
PQ0433  
genome polyprotein - human hepatitis A virus (strain S23-1) (fragment)  
C;Species: human hepatitis A virus  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C;Accession: PQ0433  
R;Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A;Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A;Reference number: PQ0427; MUID:92300330; PMID:1318940  
A;Accession: PQ0433  
A;Molecule type: mRNA  
A;Residues: 1-56 <ROB>  
A;Note: this protein is from the VP1/2A Junction region  
C;Superfamily: hepatitis A virus genome polyprotein  
C;Keywords: coat protein; core protein; polyprotein

Query Match 47.6%; Score 49; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEEL 10  
| | | | | | | |  
Db 47 QRLKYAQEEL 56

PQ0430  
genome polyprotein - human hepatitis A virus (strain S23-1) (fragment)  
C;Species: human hepatitis A virus  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C;Accession: PQ0430  
R;Robertson, B.H.; Jansen, R.W.; Khanna, B.; Totsuka, A.; Nainan, O.V.; Siegl, G.; Widel  
J. Gen. Virol. 73, 1365-1377, 1992  
A;Title: Genetic relatedness of hepatitis A virus strains recovered from different geogr  
A;Reference number: PQ0427; MUID:92300330; PMID:1318940  
A;Accession: PQ0430  
A;Molecule type: mRNA  
A;Residues: 1-56 <ROB>  
A;Note: this protein is from the VP1/2A Junction region  
C;Superfamily: hepatitis A virus genome polyprotein  
C;Keywords: coat protein; core protein; polyprotein

Query Match 47.6%; Score 49; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 0.78;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEEL 10  
| | | | | | | |  
Db 47 QRLKYAQEEL 56

RESULT 22  
S28976  
DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - human  
A;Alternate names: DNA-directed RNA polymerase B largest chain; DNA-directed RNA polymer  
C;Species: Homo sapiens (man)  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 18-Jun-1999  
C;Accession: S28976; S18986  
R;Acker, J.; Wintzerith, M.; Vigneron, M.; Keding, C.  
J. Mol. Biol. 226, 1295-1299, 1992  
A;Title: Primary structure of the second largest subunit of human RNA polymerase II (or  
A;Reference number: S28976; MUID:92389336; PMID:1518060  
A;Accession: S28976  
A;Molecule type: mRNA  
A;Residues: 1-1174 <ACK>  
A;Cross-references: EMBL:X63563; NID:936121; PIDN:CAA45124.1; PID:G36122  
C;Superfamily: DNA-directed RNA polymerase 132K polypeptide  
C;Keywords: nucleotidyltransferase; transcription

Query Match 47.6%; Score 49; DB 2; Length 1174;  
Best Local Similarity 56.2%; Pred. No. 25;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLP 16  
| | | | | | | |  
Db 334 KRIKYAKEVLOKEMLP 349

RESULT 23  
A71969  
probable histidine kinase sensor protein - Helicobacter pylori (strain J99)  
C;Species: Helicobacter pylori  
A;Variety: strain J99  
C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
C;Accession: A71969  
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
Nature 397, 176-180, 1999  
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A;Reference number: A71800; MUID:99120557; PMID:9923682  
A;Accession: A71969  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-442 <ARN>  
A;Cross-references: GB:AE001453; GB:AE001439; NID:94154651; PIDN:AAD05720.1; PID:9415465  
A;Experimental source: strain J99  
C;Genetics:  
A;Gene: jhp0151

Query Match 46.6%; Score 48; DB 2; Length 442;  
 Best Local Similarity 75.0%; Pred. No. 12;  
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 ELSNEVLPPPK 20  
 | | | | | | | | | |  
 Db 414 EADNEELPPPK 425

RESULT 24  
 A27826  
 DNA-directed RNA polymerase (EC 2.7.7.6) II second largest chain - fruit fly (Drosophila)  
 C:Species: Drosophila melanogaster  
 C:Date: 19-May-1989 #sequence\_revision 23-Nov-1991 #text\_change 21-Jul-2000  
 C:Accession: A27826; P00154  
 R:Falkenburg, D.; Dworniczak, B.; Faust, D.M.; Bautz, E.K.F.  
 J. Mol. Biol. 195, 923-937, 1987  
 A:Title: RNA polymerase II of Drosophila. Relation of its 140,000 Mr subunit to the beta  
 A:Reference number: A27826; MUID:88011299; PMID:3116266  
 A:Accession: A27826  
 A:Molecule type: DNA  
 A:Residues: 54-1176 <FAL>  
 A:Cross-references: GB:X05709; GB:M29646; NID:G5514651; PIDN:CAA29180.2; PID:G5514652  
 R:Sitzler, S.; Oldenburg, I.; Petersen, G.; Bautz, E.K.F.  
 Gene 100, 155-162, 1991  
 A:Title: Analysis of the promoter region of the housekeeping gene DmRP140 by sequence co  
 A:Reference number: P00154; MUID:91276237; PMID:1905256  
 A:Accession: P00154  
 A:Molecule type: mRNA  
 A:Residues: 1-69 <SIT>  
 A:Cross-references: GB:M62972; NID:G157263; PIDN:AAA28476.1; PID:G157267  
 A:Experimental source: embryo  
 C:Genetics:  
 A:Gene: DmRP140  
 A:Cross-references: FlyBase:FBgn0003276  
 A:Introns: 8/1  
 C:Superfamily: DNA-directed RNA polymerase 132K polypeptide  
 C:Keywords: DNA binding; nucleotidyltransferase; nucleus; transcription; zinc finger

Query Match 46.6%; Score 48; DB 2; Length 1176;  
 Best Local Similarity 56.2%; Pred. No. 36;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRLKYAQELSNEVL 16  
 | | | | | | | | | |  
 Db 335 KRIKYAKEILQKEMLP 350

RESULT 25  
 S65068  
 DNA-directed RNA polymerase (EC 2.7.7.6) II second largest chain - tomato  
 C:Species: Lycopersicon esculentum (tomato)  
 C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 18-Jun-1999  
 C:Accession: S65068  
 R:Wartlow, D.; Symons, R.H.  
 Plant Mol. Biol. 30, 337-342, 1996  
 A:Title: Sequence analysis of the second largest subunit of tomato RNA polymerase II.  
 A:Reference number: S65068; MUID:96178872; PMID:8616257  
 A:Accession: S65068  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-1191 <WAR>  
 A:Cross-references: EMBL:U28403; NID:G1049067; PIDN:AAC49273.1; PID:G1049068  
 C:Genetics:  
 A:Gene: RPB2  
 C:Superfamily: DNA-directed RNA polymerase 132K polypeptide  
 C:Keywords: DNA binding; nucleotidyltransferase; nucleus; transcription

Query Match 46.6%; Score 48; DB 2; Length 1191;  
 Best Local Similarity 56.2%; Pred. No. 36;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRLKYAQELSNEVL 16

Db 338 KRIKYAKEILQKEMLP 353  
 | | | | | | | | | |  
 RESULT 26  
 LLBY  
 actin-binding protein ABP1 - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein YCR088w  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 31-Mar-1991 #sequence\_revision 31-Dec-1993 #text\_change 05-Nov-1999  
 C:Accession: S19503; S19767; S07608  
 R:Dusterhoft, A.; Erdmann, D.; Hagemann, J.; Philippesen, P.; Schweitzer, B.; Spiegelberg,  
 submitted to the Protein Sequence Database, March 1992  
 A:Reference number: S19351  
 A:Accession: S19503  
 A:Molecule type: DNA  
 A:Residues: 1-476 <DUS>  
 A:Cross-references: EMBL:X59720; GSPDB:GN00003; MIPS:YCR088w  
 R:Frontali, L.; Grisanti, P.  
 submitted to the Protein Sequence Database, March 1992  
 A:Reference number: S19504  
 A:Accession: S19767  
 A:Molecule type: DNA  
 A:Residues: 14-592 <PRO>  
 A:Cross-references: EMBL:X59720; GSPDB:GN00003; MIPS:YCR088w  
 R:Drubin, D.G.; Muhlolland, J.; Zhu, Z.; Botstein, D.  
 Nature 343, 288-290, 1990  
 A:Title: Homology of a yeast actin-binding protein to signal transduction proteins and m  
 A:Reference number: S07608; MUID:90136906; PMID:2405279  
 A:Accession: S07608  
 A:Molecule type: DNA  
 A:Residues: 1-57, 'S', 59-311, 'I', 313-592 <DRU>  
 A:Cross-references: EMBL:X51780; NID:G3321; PIDN:CAA36075.1; PID:G3322  
 C:Genetics:  
 A:Gene: SGD:ABP1; MIPS:YCR088w  
 A:Cross-references: SGD:S0000684; MIPS:YCR088w  
 A:Map position: 3R  
 C:Superfamily: actin-binding protein ABP1; SH3 homology  
 C:Keywords: actin binding; duplication  
 F:158-180.378-401/Region: duplication  
 F:539-588/Domain: SH3 homology <SH3>

Query Match 45.8%; Score 47; DB 1; Length 592;  
 Best Local Similarity 50.0%; Pred. No. 24;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 7 QEELSNEVLPPPK 20  
 | | | | | | | | | |  
 Db 369 EEMENKFPAPPKK 382

RESULT 27  
 T29757  
 protein UNC-89 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-Dec-1999  
 C:Accession: T29757  
 R:Du, Z.; Le, T.T.; Wilson, R.  
 submitted to the EMBL Data Library, May 1997  
 A:Description: The sequence of C. elegans cosmid C09D1.  
 A:Reference number: Z20679  
 A:Accession: T29757  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-6642 <DUZ>  
 A:Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89  
 A:Experimental source: strain Bristol N2; clone C09D1  
 C:Genetics:  
 A:Gene: CESP:unc-89  
 A:Map position: 1  
 A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 61  
 /3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1



Query Match 45.6%; Score 47; DB 2; Length 6642;  
Best Local Similarity 45.0%; Pred. No. 3.7e+02;  
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 1353 RRVSAEEELPKXEVDSORK 1372  
RESULT 28  
156893  
transcription factor AP-4 - human  
C;Species: Homo sapiens (man)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
C;Accession: I56893; A36394  
R;Ou, S.H.; Garcia-Martinez, L.F.; Paulssen, E.J.; Gaynor, R.B.  
J. Virol. 68, 7188-7199, 1994  
A;Title: Role of flanking E box motifs in human immunodeficiency virus type 1 TATA element  
A;Reference number: I56893; MUID:95018629; PMID:7933101  
A;Accession: I56893  
A;Status: Preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-338 <RES>  
A;Cross-references: GB:S73885; NID:G693848; PIDN:AAB32235.1; PID:G693849  
R;Hu, Y.F.; Luescher, B.; Admon, A.; Mermoud, N.; Tjian, R.  
Genes Dev. 4, 1741-1752, 1990  
A;Title: Transcription factor AP-4 contains multiple dimerization domains that regulate  
A;Reference number: A36394; MUID:91065520; PMID:2123466  
A;Accession: A36394  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 18-338 <HUA>  
A;Cross-references: GB:X57435; NID:G37059; PIDN:CAA40683.1; PID:G37060  
C;Genetics:  
A;Gene: GDB:TFAP4; AP-4  
A;Cross-references: GDB:455279; OMIM:600743  
A;Map position: 16p13-16p13  
C;Keywords: DNA binding; transcription regulation  
Query Match 44.7%; Score 46; DB 2; Length 338;  
Best Local Similarity 38.9%; Pred. No. 18;  
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPP 18  
Db 210 EKLREQQQLRTQLPPPP 227  
RESULT 29  
FWPUB  
11S globulin beta subunit precursor - cucurbita cv. Kurokawa Amakuri  
N;Alternate names: globulin IV beta subunit  
C;Species: Cucurbita cv. Kurokawa Amakuri  
A;Variety: cv. Kurokawa Amakuri Nankin  
C;Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 23-Sep-2002  
C;Accession: S00366; S09480  
R;Hayashi, M.; Mori, H.; Nishimura, M.; Akazawa, T.; Hara-Nishimura, I.  
Eur. J. Biochem. 172, 627-632, 1988  
A;Title: Nucleotide sequence of cloned cDNA coding for pumpkin 11-S globulin beta subunit  
A;Reference number: S00366; MUID:88166744; PMID:2450746  
A;Accession: S00366  
A;Molecule type: mRNA  
A;Residues: 1-480 <HAY>  
A;Cross-references: EMBL:M36407; NID:G167491; PIDN:AAA33110.1; PID:G167492  
A;Experimental source: Cucurbita sp. cv. Kurokawa Amakuri Nankin; cotyledon mRNA  
R;Ohmiya, M.; Hara, I.; Matsubara, H.  
Plant Cell Physiol. 21, 157-167, 1980  
A;Title: Pumpkin (Cucurbita sp.) seed globulin IV. Terminal sequences of the acidic and  
A;Reference number: S09066  
A;Accession: S09480  
A;Molecule type: protein  
A;Residues: 22-26,'E',28-29,'S',297-302 <OHM>  
A;Experimental source: Cucurbita sp. hybrid, Tetsukabuto-Nankin

A;Note: the carboxyl end of the gamma' chain is Leu or Val; the carboxyl end of the delta  
C;Complex: heterohexamer; hexamer of alpha and beta subunits; beta subunit composed of d  
C;Superfamily: glycinin  
C;Keywords: pyroglutamic acid; seed; storage protein  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-274/Product: (or 22-272) 11S globulin gamma' chain #status experimental <GCH>  
F;297-479/Product: (or 297-478) 11S globulin delta-2 chain #status experimental <DCH>  
F;124-303/Disulfide bonds: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen  
Query Match 44.7%; Score 46; DB 1; Length 480;  
Best Local Similarity 57.9%; Pred. No. 27;  
Matches 11; Conservative 2; Mismatches 4; Indels 2; Gaps 1;  
QY 1 QRLKYAQEELSNEVLPPPP 19  
Db 458 QRLKYQCEM-RVLSPPGR 474  
RESULT 30  
T01053  
hypothetical protein YUP8H12R.36 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 24-Nov-1999  
C;Accession: T01053  
R;Theologis, A.; Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; Kwan  
Oefner, P.; Davis, R.W.  
submitted to the EMBL Data Library, May 1998  
A;Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.  
A;Reference number: Z14227  
A;Accession: T01053  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-483 <THE>  
A;Cross-references: EMBL:AC002986; NID:G2494106; PID:G3152596; GSPDB:GN00059; ATSP:YUP8H  
C;Genetics:  
A;Gene: ATSP:YUP8H12R.36  
A;Map position: 1  
A;Introns: 52/3; 96/1; 141/3; 154/3; 167/2; 208/3; 230/3; 243/3; 259/3; 279/3; 296/3; 31  
C;Superfamily: Arabidopsis thaliana hypothetical protein YUP8H12R.36  
Query Match 44.7%; Score 46; DB 2; Length 483;  
Best Local Similarity 50.0%; Pred. No. 27;  
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 40 QLLRIASVEVDNEVAPVPSK 59  
Search completed: May 11, 2004, 13:41:18  
Job time : 12.75 secs

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:24:26 ; Search time 6.75 Seconds  
(without alignments)  
154.282 Million cell updates/sec

Title: US-09-171-432A-41

Perfect score: 103

Sequence: 1 QRLKYAQEELSNEVLPPPK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	103	100.0	341	1	POLG_HPAV1	P13672 hepatitis a
2	103	100.0	808	1	POLG_HPAV1	Q02381 hepatitis a
3	103	100.0	852	1	POLG_HPAV1	P05442 hepatitis a
4	103	100.0	2226	1	POLG_HPAV2	P26580 hepatitis a
5	103	100.0	2226	1	POLG_HPAV2	P26581 hepatitis a
6	103	100.0	2226	1	POLG_HPAV8	P26582 hepatitis a
7	103	100.0	2227	1	POLG_HPAV8	P08617 hepatitis a
8	103	100.0	2227	1	POLG_HPAV1	P05441 hepatitis a
9	103	100.0	2227	1	POLG_HPAV1	P13901 hepatitis a
10	97	94.2	2230	1	POLG_HPAV5	P14553 simian hepa
11	94	91.3	839	1	POLG_HPAV1	P31788 simian hepa
12	52	50.5	1193	1	RPB2_CAEEL	Q10578 caenorhabdi
13	49.5	48.1	443	1	HSJU_VIBCH	Q9knq7 vibrio chol
14	49.5	48.1	443	1	HSJU_VIBCH	Q8dcp4 vibrio vuln
15	49	47.6	1174	1	RPB2_HUMAN	P30876 homo sapien
16	48	46.6	636	1	NAF1_HUMAN	Q15025 homo sapien
17	48	46.6	1176	1	RPB2_DROME	P08266 drosophila
18	48	46.6	1191	1	RPB2_LYCES	Q42877 lycopersico
19	47	45.6	592	1	ABP1_YEAST	P15891 saccharomyc
20	47	45.6	6832	1	UN89_CAEEL	Q01761 caenorhabdi
21	46	44.7	338	1	TAP4_HUMAN	Q01664 homo sapien
22	46	44.7	480	1	11S8_CUCMA	P13744 cucurbita m
23	45	43.7	733	1	HEXA_BLADI	Q17127 blaberus di
24	45	43.7	1188	1	RPB2_ARATH	P38420 arabidopsis
25	44	42.7	303	1	PM33_YEAST	Q12326 saccharomyc
26	44	42.7	443	1	HSJU_YERPE	Q82j55 yersinia pe
27	44	42.7	510	1	GPW1_ANTSP	Q06464 antithamio
28	44	42.7	914	1	ORC1_YEAST	P54784 saccharomyc
29	43.5	42.2	3678	1	DMD_MOUSE	P11531 mus musculu
30	43.5	42.2	3680	1	DMD_CANFA	Q97592 canis famil
31	43.5	42.2	3685	1	DMD_HUMAN	P11532 homo sapien
32	43	41.7	120	1	YLOU_BACSU	Q34318 bacillus su
33	43	41.7	268	1	YHC6_YEAST	P38740 saccharomyc

RESULT 1

#### ALIGNMENTS

34	43	41.7	574	1	HEMA_INBMD	P03461 influenza b
35	43	41.7	576	1	HEMA_INBUS	P09766 influenza b
36	43	41.7	578	1	HEMA_INBME	P09765 influenza b
37	43	41.7	578	1	HEMA_INBVI	P09767 influenza b
38	43	41.7	583	1	HEMA_INBEN	P10757 influenza b
39	43	41.7	583	1	HEMA_INBOR	P03464 influenza b
40	43	41.7	583	1	HEMA_INBSI	P03463 influenza b
41	43	41.7	585	1	HEMA_INBE	P17504 influenza b
42	43	41.7	585	1	HEMA_INBVK	P22092 influenza b
43	43	41.7	1146	1	AS10_YEAST	P48361 saccharomyc
44	43	41.7	1210	1	RPB2_SCHPO	Q02061 schizosacch
45	43	41.7	2696	1	NSD1_HUMAN	Q06173 homo sapien
46	42.5	41.3	2278	1	FAB1_YEAST	P34756 saccharomyc
47	42	40.8	214	1	DEV5_MYXXA	Q07766 myxococcu
48	42	40.8	276	1	YJIC_ECOLI	P39374 escherichia
49	42	40.8	445	1	ML64_HUMAN	Q14849 homo sapien
50	42	40.8	554	1	PEX2_HUMAN	P51021 podospora a
51	42	40.8	1005	1	DPOL_VARV	P33793 variola vir
52	42	40.8	1006	1	DPOL_VACCC	P20509 vaccinia vi
53	42	40.8	1006	1	DPOL_VACCV	P06856 vaccinia vi
54	42	40.8	1066	1	AOSL_PLEHO	Q16025 plexaura ho
55	42	40.8	1270	1	DXH9_HUMAN	Q08211 homo sapien
56	42	40.8	1773	1	DIP2_DROME	Q9W089 drosophila
57	41.5	40.3	348	1	OMH1_PASMU	Q9cmn9 pasteurella
58	41.5	40.3	498	1	PM2_DROME	P31369 drosophila
59	41	39.8	181	1	ATPF_STRLI	P50013 streptomyce
60	41	39.8	258	1	RT15_MOUSE	Q9dc71 mus musculu
61	41	39.8	321	1	PE28_ARATH	Q9s677 arabidopsis
62	41	39.8	349	1	F161_RHIME	Q9exv4 rhizobium m
63	41	39.8	349	1	F162_RHIME	P56886 rhizobium m
64	41	39.8	451	1	CSSS_BACSU	Q32193 bacillus su
65	41	39.8	460	1	ANT3_HUMAN	Q9Y5c1 homo sapien
66	41	39.8	500	1	CISA_BACSU	P17867 bacillus su
67	41	39.8	512	1	FLOH_MOUSE	P41438 mus musculu
68	41	39.8	519	1	TRPE_SERMA	P00897 serratia ma
69	41	39.8	758	1	YP58_YEAST	Q99299 saccharomyc
70	41	39.8	789	1	ATX1_RAT	Q63540 rattus norv
71	41	39.8	816	1	ATX1_HUMAN	P54253 homo sapien
72	41	39.8	1719	1	PRD2_HUMAN	Q13029 homo sapien
73	41	39.8	1966	1	CCAF_HUMAN	O60840 homo sapien
74	41	39.8	1985	1	CCAF_MOUSE	Q9J1s7 mus musculu
75	40.5	39.3	327	1	HUNE_MANSE	Q25514 manduca sex
76	40	38.8	319	1	DBP8_YEAST	P38637 saccharomyc
77	40	38.8	443	1	Y314_MYCGE	Q49415 mycoplasma
78	40	38.8	584	1	HCM1_YEAST	P25364 saccharomyc
79	40	38.8	624	1	ARLY_VIBVU	Q8dcm9 vibrio vuln
80	40	38.8	681	1	COG2_CAEEL	Q21444 caenorhabdi
81	40	38.8	802	1	OPH1_MOUSE	Q99731 mus musculu
82	40	38.8	853	1	LOX4_SOYBN	P38417 glycine max
83	40	38.8	935	1	MYCD_MOUSE	Q8vims mus musculu
84	40	38.8	938	1	MYCD_RAT	Q8vsi7 rattus norv
85	40	38.8	1097	1	CCT_DROME	Q96433 drosophila
86	40	38.8	1197	1	SMC2_HUMAN	O95347 homo sapien
87	40	38.8	1229	1	SIP3_YEAST	P38717 saccharomyc
88	40	38.8	1380	1	DXK3_MOUSE	O70133 mus musculu
89	40	38.8	1398	1	TRC3_MOUSE	O90900 m nuclear r
90	40	38.8	2032	1	NPQ3_CAEEL	Q93971 caenorhabdi
91	40	38.8	2390	1	SPCP_HUMAN	O15020 homo sapien
92	40	38.8	2476	1	ATRX_MOUSE	O61687 mus musculu
93	39.5	38.3	482	1	YSR2_CAEEL	Q09950 caenorhabdi
94	39.5	38.3	530	1	GME2_HUMAN	Q9ukd1 homo sapien
95	39.5	38.3	702	1	CRN_DROME	P17886 drosophila
96	39.5	38.3	867	1	NIA_ASPNG	P36858 aspergillus
97	39.5	38.3	908	1	ATME_SALTU	P22036 salmonella
98	39.5	38.3	931	1	MCW4_SCHPO	P29458 schizosacch
99	39.5	38.3	1589	1	PHP_DROME	P39769 drosophila
100	39.5	38.3	2241	1	TEGU_HCNVA	P16785 human cytom

```
POLG_HPAV1 STANDARD; PRT; 341 AA.
ID POLG_HPAV1 STANDARD; PRT; 341 AA.
AC P13672;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Coat protein VP3 (1C); Coat protein VP1 (1D); Core protein P2A] (Fragment).
DS Hepatitis A virus (strain LCDC-1).
OS Hepatitis A virus (ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Hepatovirus).
OC Hepatovirus.
OX NCBI_TaxID=12093;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89263805; PubMed=2542903;
RA Aronov A.P., Lau P., Chaudhary R.;
RT "Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A virus (HAV).";
RL Nucleic Acids Res. 17:3594-3594(1989).
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units, each of which is composed of one copy each of proteins VP1, VP2, VP3, and VP4.
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; X14666; CAA32794.1; -.
CC PIR; S04137.
CC InterPro; IPR008975; Viral cap coat.
CC Polyprotein; Coat protein; Core protein.
FT NON_TER 1 1 COAT PROTEIN VP3.
FT CHAIN <1 1 COAT PROTEIN VP1.
FT CHAIN 2 340 COAT PROTEIN P2A.
FT CHAIN 341 >341 CORE PROTEIN P2A.
FT NON_TER 341 341
SQ SEQUENCE 341 AA; 38003 MW; 066918289BF126D5 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
DB 314 QRLKYAQEELSNEVLPPPRK 333

RESULT 2
POLG_HPAVG STANDARD; PRT; 808 AA.
ID POLG_HPAVG STANDARD; PRT; 808 AA.
AC Q02381;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein P2A] (Fragment).
DS Hepatitis A virus (strain CA76).
OS Hepatitis A virus (ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Hepatovirus).
OC Hepatovirus.
OX NCBI_TaxID=31706;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92260183; PubMed=1316423;
RA Khanna B., Spelbring J.E., Innis B.L., Robertson B.H.;
RT "Characterization of a genetic variant of human hepatitis A virus.";
RL J. Med. Virol. 36:118-124(1992).
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units, each of which is composed of one copy each of proteins VP1, VP2,
```

```
VP3, and VP4.
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; M66695; AAA45477.1; -.
CC InterPro; IPR008975; Viral cap coat.
CC Polyprotein; Coat protein; Core protein.
FT NON_TER 1 1 COAT PROTEIN VP4 (P1A).
FT CHAIN <1 2 COAT PROTEIN VP2 (P1B).
FT CHAIN 3 223 COAT PROTEIN VP3 (P1C).
FT CHAIN 224 470 COAT PROTEIN VP1 (P1D).
FT CHAIN 471 770 COAT PROTEIN P2A.
FT CHAIN 771 >808
FT NON_TER 808 808
SQ SEQUENCE 808 AA; 90632 MW; D80CE7E57A479C12 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 808;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
DB 789 QRLKYAQEELSNEVLPPPRK 808

RESULT 3
POLG_HPAVC STANDARD; PRT; 852 AA.
ID POLG_HPAVC STANDARD; PRT; 852 AA.
AC P06442; Q83741; Q83742;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein P2A] (Fragment).
DS Hepatitis A virus (strain CR326).
OS Hepatitis A virus (ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Hepatovirus).
OC Hepatovirus.
OX NCBI_TaxID=12097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85185648; PubMed=2985793;
RA Linemeyer D.L., Menke J.G., Martin-Gallardo A., Hughes J.V., Young A., Mitra S.W.;
RT "Molecular cloning and partial sequencing of hepatitis A viral cDNA.";
RL J. Virol. 54:247-255(1985).
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units, each of which is composed of one copy each of proteins VP1, VP2, VP3, and VP4.
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; M10033; AAA45470.1; -.
CC PIR; A03904; GNHYHA.
CC InterPro; IPR008975; Viral cap coat.
CC Polyprotein; Coat protein; Core protein.
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).
```

FT CHAIN 492 836 COAT PROTEIN VP1 (P1D).  
 FT CHAIN 837 >852 CORE PROTEIN P2A.  
 FT NON TER 852 852  
 SQ SEQUENCE 852 AA; 95563 MW; 73D3ED0AD532820E CRC64;  
 Query Match 100.0%; Score 103; DB 1; Length 852;  
 Best Local Similarity 100.0%; Pred. No. 3e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPK 20  
 |||||  
 DB 810 QRLKYAQEELSNEVLPPPK 829

RESULT 4  
 POLG\_HPAV2  
 ID POLG\_HPAV2 STANDARD; PRT; 2226 AA.  
 AC P25580;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
 DE P3D (EC 2.7.7.48).  
 OS Hepatitis A virus (strain 24a).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12094;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91162758; PubMed=1705995;  
 RX Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
 RA Cromeans T., Jansen R.W.;  
 RA "Antigenic and genetic variation in cytopathic hepatitis A virus  
 RT variants arising during persistent infection: evidence for genetic  
 RT recombination.";  
 RL J. Virol. 65:2056-2065(1991).  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 (RNA)(N).  
 CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M59810; AAA45468.1; -.  
 CC MEROPS; C03.005; -.  
 DR InterPro; IPR004004; Calici\_pol\_hel.  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR000605; RNA helicase.  
 DR InterPro; IPR007095; RNA pol DS PS.  
 DR InterPro; IPR001205; RNA pol P3D.  
 DR InterPro; IPR007094; RNA pol\_Psvir.  
 DR InterPro; IPR008975; Viral\_cap\_coat.  
 DR Pfam; PF00680; RNA dep\_RNA\_pol; 1.  
 DR Pfam; PF00910; RNA helicase; 1.  
 DR PRINTS; PR00918; CALICVIRUSNS.  
 KW Polyprotein; Coat protein; Core protein; Transferase;  
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
 FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
 FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
 FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
 FT CHAIN 492 794 COAT PROTEIN VP1 (P1D).  
 FT CHAIN 795 900 CORE PROTEIN P2A.  
 FT CHAIN 901 1087 CORE PROTEIN P2B.

FT CHAIN 1088 1422 CORE PROTEIN P2C.  
 FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
 FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
 FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
 FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
 SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;  
 Query Match 100.0%; Score 103; DB 1; Length 2226;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPK 20  
 |||||  
 DB 810 QRLKYAQEELSNEVLPPPK 829

RESULT 5  
 POLG\_HPAV4  
 ID POLG\_HPAV4 STANDARD; PRT; 2226 AA.  
 AC P26581;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
 DE P3D (EC 2.7.7.48).  
 OS Hepatitis A virus (strain 43c).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12095;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91162758; PubMed=1705995;  
 RX Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
 RA Cromeans T., Jansen R.W.;  
 RA "Antigenic and genetic variation in cytopathic hepatitis A virus  
 RT variants arising during persistent infection: evidence for genetic  
 RT recombination.";  
 RL J. Virol. 65:2056-2065(1991).  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 (RNA)(N).  
 CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M59809; AAA45469.1; -.  
 CC MEROPS; C03.005; -.  
 DR InterPro; IPR004004; Calici\_pol\_hel.  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR000605; RNA helicase.  
 DR InterPro; IPR007095; RNA pol DS PS.  
 DR InterPro; IPR001205; RNA pol P3D.  
 DR InterPro; IPR007094; RNA pol\_Psvir.  
 DR InterPro; IPR008975; Viral\_cap\_coat.  
 DR Pfam; PF00680; RNA dep\_RNA\_pol; 1.  
 DR Pfam; PF00910; RNA helicase; 1.  
 DR PRINTS; PR00918; CALICVIRUSNS.  
 KW Polyprotein; Coat protein; Core protein; Transferase;  
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
 FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
 FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
 FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
 FT CHAIN 492 794 COAT PROTEIN VP1 (P1D).

```

FT CHAIN 795 900 CORE PROTEIN P2A.
FT CHAIN 901 1087 CORE PROTEIN P2B.
FT CHAIN 1088 1422 CORE PROTEIN P2C.
FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.
FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.
FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.
FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.
SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B09BF75 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 2226;
Best Local Similarity 100.0%; Pred. No. 8,8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20
    |||||
    |||||
DB 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 6
POLG_HPAV8
ID POLG_HPAV8 STANDARD; PRT; 2226 AA.
AC P26582;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
DE P3D (EC 2.7.7.48)].
DE Hpatitis A virus (strain 18F).
OS Hpatitis A virus (strain 18F).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
NCBI_TaxID=12096;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=91162758; PubMed=1705995;
RX Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstein S.M.,
RA Cromeans T., Jansen R.W.;
RT "Antigenic and genetic variation in cytopathic hepatitis A virus
RT variants arising during persistent infection: evidence for genetic
RT recombination."
RT J. Virol. 68:2056-2065(1991).
RL
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
CC each of which is composed of one copy each of proteins VP1, VP2,
CC VP3, and VP4.
CC
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; M59808; AAA5467.1; --
CC PDB; 1QAV; 15-MAY-00.
CC
CC MEROPS; C03.005; -.
CC
CC InterPro; IPR004004; Calici_pol_hel.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR000605; RNA_helicase.
CC InterPro; IPR007095; RNA_pol_DS_PS.
CC InterPro; IPR001205; RNA_pol_P3D.
CC InterPro; IPR007094; RNA_pol_PSVir.
CC InterPro; IPR008975; Viral_cap_coat.
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Pfam; PF00910; RNA_helicase; 1.
CC PRINTS; PR00918; CALICIVIRUSN.
CC
CC Polyprotein; Coat protein; Core protein; Transferase;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.
CC CHAIN 1 23 COAT PROTEIN VP4 (PIA).

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC CC the European Bioinformatics Institute. There are no restrictions on its  
CC CC use by non-profit institutions as long as its content is in no way  
CC CC modified and this statement is not removed. Usage by and for commercial  
CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M14114; AAA45475.1; --  
DR EMBL; M14707; AAA45465.1; --  
DR EMBL; M14707; AAA45466.1; ALT INIT.  
DR EMBL; M16632; AAA45471.1; --  
DR PIR; A03905; A03905.  
DR PIR; A25981; GNNYHM.  
DR PIR; A94149; GNNYMK.  
DR PDB; 1HAV; 23-DEC-96.  
DR MEROPS; C03.005; --  
DR InterPro; IPR004004; Calici\_pol\_hel.  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR006005; RNA\_helicase.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICIVIRUS.  
DR Polyprotein; Coat protein; Core protein; Transferase;  
RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 831  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1087  
FT CHAIN 1088 1422  
FT CHAIN 1423 1496  
FT CHAIN 1497 1519  
FT CHAIN 1520 1738  
FT CHAIN 1739 2227  
FT VARIANT 77 77  
FT VARIANT 764 764  
FT VARIANT 821 821  
FT VARIANT 1052 1052  
FT VARIANT 1062 1062  
FT VARIANT 1118 1118  
FT VARIANT 1151 1151  
FT VARIANT 1163 1163  
FT VARIANT 1277 1277  
FT VARIANT 1500 1500  
FT VARIANT 1805 1805  
FT VARIANT 1930 1930  
SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7AEB740A6 CRC64;

Query Match 100.0%; Score 103; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
|||  
DB 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 8

POLG HPAVL  
ID POLG HPAVL STANDARD; PRT; 2227 AA.  
AC P05441.  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein [contains: Coat proteins VP1 TO VP4; Core proteins  
DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase

DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain IA).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85190549; PubMed=2986127;  
RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,  
RA Merryweather J., van Nest G., Dina D.;  
RT "Primary structure and gene organization of human hepatitis A virus.";  
RT Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC {RNA} (N).  
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; K02990; AAA45472.1; --  
DR PIR; A03903; GNNYHR.  
DR MEROPS; C03.005; --  
DR InterPro; IPR004004; Calici\_pol\_hel.  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR006005; RNA\_helicase.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICIVIRUS.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23  
FT CHAIN 24 245  
FT CHAIN 246 491  
FT CHAIN 492 836  
FT CHAIN 837 980  
FT CHAIN 981 1076  
FT CHAIN 1077 1422  
FT CHAIN 1423 1484  
FT CHAIN 1485 1507  
FT CHAIN 1508 1678  
FT CHAIN 1679 2227  
SQ SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;

Query Match 100.0%; Score 103; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 8.8e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
|||  
DB 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 9

POLG HPAVM  
ID POLG HPAVM STANDARD; PRT; 2227 AA.  
AC P13901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;  
AC Q81090; Q81091; Q81092; Q81093;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)



Matches	18;	Conservative	1;	Mismatches	1;	Indels	0;	Gaps	0;
QY	1	ORLKYAQEELSNEVLPPPRK	20						
DB	814	ORLKYAQEELSNEILPPPRK	833						
RESULT 11									
POLG	HPAVT	STANDARD;	PRT;	839	AA.				
ID									
AC	P31788;								
DT	01-JUL-1993	(Rel. 26, Created)							
DT	01-JUL-1993	(Rel. 26, Last sequence update)							
DT	16-OCT-2001	(Rel. 40, Last annotation update)							
DE	Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein P2A] (Fragment)								
DE	Simian hepatitis A virus (strain CV-145).								
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;								
OC	Hepatovirus.								
OC	NCBI_TaxID=31707;								
OX	[1]								
RN	SEQUENCE FROM N.A.								
RP	MEDLINE=91311421; PubMed=1649902;								
RA	Nainan O.V., Margolis H.S., Robertson B.H., Balayan M., Brinton M.A.;								
RX	"sequence analysis of a new hepatitis A virus naturally infecting								
RT	synomolgus macaques (Macaca fascicularis).";								
RL	J. Gen. Virol. 72:1685-1689(1991).								
RL	-!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,								
CC	each of which is composed of one copy each of proteins VP1, VP2,								
CC	VP3, and VP4.								
CC	-!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.								
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -								
CC	the European Bioinformatics Institute. There are no restrictions on its								
CC	use by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>								
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).								
CC									
CC	EMBL; M59286; AAA45473.1; --								
DR	PIR; JQ1180; GNNYS2.								
DR	InterPro; IPR008975; Viral_cap_coat.								
KW	Polyprotein; Coat protein; Core protein.								
FT	CHAIN 1 23	COAT PROTEIN VP4 (PIA).							
FT	CHAIN 24 245	COAT PROTEIN VP2 (PIB).							
FT	CHAIN 246 491	COAT PROTEIN VP3 (PIC).							
FT	CHAIN 492 ?	COAT PROTEIN VP1 (PID).							
FT	CHAIN ? >839	CORE PROTEIN P2A.							
FT	NON TER 839								
FT	SEQUENCE 839 AA; 93825 MW; 2ACCC4BD1E192DBC CRC64;								
Query Match									
Best Local Similarity	91.3%;	Score 94;	DB 1;	Length 839;					
Matches 17;	Conservative	2;	Mismatches	1;	Indels	0;	Gaps	0;	

```

QY      1  ORLKYAQEELSNEVLPPPRK  20
      |||:|||||:|||||
      |||:|||||:|||||
Db      809  QRFKYAREELSNEILPPPRK  828

RESULT 12
RPB2 CAEL
ID   RPB2 CAEL          STANDARD;          PRT; 1193 AA.
AC   Q10578;
DT   01-OCT-1996 (Rel. 34, Created)
DD   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   DNA-directed RNA polymerase II second largest subunit (EC 2.7.7.6)
DE   (RNA polymerase I subunit 2).
DE   C26E6.4.
GN   Caeenorhabditis elegans.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

```

Rhabditidae; Peloderinae; Caenorhabditis.  
NCBI\_taxid=6239;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=Bristol N2;  
Fulton L.;  
Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
[2]  
SEQUENCE OF 200-1058 FROM N.A.  
MEDLINE=95041334; PubMed=7953533;  
Sidow A., Thomas W.K.;  
"A molecular evolutionary framework for eukaryotic model organisms.";  
Curr. Biol. 4:596-603(1994).  
-!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
of DNA into RNA using the four ribonucleoside triphosphates as  
substrates.  
-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
(RNA) (N).  
-!- SUBUNIT: Each class of RNA polymerase is assembled from 9 to 14  
different polypeptides. This subunit is the second largest  
component of RNA polymerase II.  
-!- SUBCELLULAR LOCATION: Nuclear.  
-!- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases are  
found in eukaryotic nuclei: polymerase I for the ribosomal RNA  
precursor, polymerase II for the mRNA precursor, and polymerase  
III for 5S and rRNA genes.  
-!- SIMILARITY: Belongs to the RNA polymerase beta chain family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

EMBL; U13875;	AA221158.1; -.
DR	DR
EMBL; U10333;	AAA50224.1; -.
DR	DR
PIR; E88445;	E88445.
DR	DR
PIR; T43701;	T43701.
DR	DR
WormPep; C26E6.4;	CE01162.
DR	DR
InterPro; IPR007121;	RNA_pol_B.
DR	DR
InterPro; IPR007644;	RNA_pol_Rpb2_1.
DR	DR
InterPro; IPR007642;	RNA_pol_Rpb2_2.
DR	DR
InterPro; IPR007645;	RNA_pol_Rpb2_3.
DR	DR
InterPro; IPR007646;	RNA_pol_Rpb2_4.
DR	DR
InterPro; IPR007647;	RNA_pol_Rpb2_5.
DR	DR
InterPro; IPR007120;	RNA_pol_Rpb2_6.
DR	DR
InterPro; IPR007641;	RNA_pol_Rpb2_7.
DR	DR
Pfam; PF04563;	RNA_pol_Rpb2_1; 1.
DR	DR
Pfam; PF04561;	RNA_pol_Rpb2_2; 1.
DR	DR
Pfam; PF04565;	RNA_pol_Rpb2_3; 1.
DR	DR
Pfam; PF04566;	RNA_pol_Rpb2_4; 1.
DR	DR
Pfam; PF04567;	RNA_pol_Rpb2_5; 1.
DR	DR
Pfam; PF00562;	RNA_pol_Rpb2_6; 1.
DR	DR
Pfam; PF04560;	RNA_pol_Rpb2_7; 1.
DR	DR
PROSITE; PS01166;	RNA_POL_BETA; 1.
DR	DR
Transferase; DNA-directed RNA polymerase;	Transcription; Zinc;
KW	KW
Zinc-finger; Nuclear protein.	
ZN FING	1125 1146
C4-TYPE (POTENTIAL).	
FT	1193 AA; 134904 MW; B8A85B74E9C7EEB CRC64;
SO	SEQUENCE

```

Query Match      50.5%; Score 52; DB 1; Length 1193;
Best Local Similarity 62.5%; Pred. No. 4.7;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1  ORLKVAQEELSNEVLP 16
      ||:|:|:|:|:|:|:|:|
Db      340  QRIKYAREILQKELLP 355

RESULT 13
HSLU VIBCH

```



ID HSLU\_VIBCH STANDARD; PRT; 443 AA.  
 AC Q9K07;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE ATP-dependent hsl protease ATP-binding subunit hslu.  
 GN HSLU OR VC2674.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=El Tor N16961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heideberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RL Nature 406:477-483(2000).  
 CC -1- FUNCTION: Chaperone subunit of a proteasome-like degradation  
 CC complex (By similarity).  
 CC -1- SUBUNIT: A double ring-shaped homohexameric of hslu is capped on  
 CC each side by a ring-shaped hslu homohexameric (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: Belongs to the clpX chaperone family. Hslu subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE004333; AAF95815.1; -.  
 DR PIR: E82046; E82046.  
 DR HSPF: P32168; IDO2.  
 DR TIGR: VC2674; -.  
 DR HAMAP: MF 00249; -; 1.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003959; AAA\_ATPase\_central.  
 DR InterPro: IPR004491; Hsp\_HsluV.  
 DR Pfam: PF00004; AAA; 1.  
 DR SMART: SM00382; AAA; 1.  
 DR TIGRFAMs: TIGR00390; hslu; 1.  
 DR KW Chaperone; ATP-binding; Complete proteome.  
 FT NP\_BIND 57 64 ATP (POTENTIAL).  
 SQ SEQUENCE 443 AA; 49900 MW; DA13E82FA6A38F CRC64;  
 Query Match 48.1%; Score 49.5; DB 1; Length 443;  
 Best Local Similarity 41.7%; Pred. No. 3.9;  
 Matches 10; Conservative 6; Mismatches 3; Indels 5; Gaps 1;  
 Qy 1 ORLKYAQBELSNE-----VLPPPP 19  
 Db 117 EKVFRAEELAEERVLDAALLPPPP 140  
 RESULT 14  
 HSLU\_VIBU  
 ID HSLU\_VIBU STANDARD; PRT; 443 AA.  
 AC Q8DCP4;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE ATP-dependent hsl protease ATP-binding subunit hslu.  
 GN HSLU OR VV11355.

OS Vibrio vulnificus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Choy H.E.;  
 RA "Complete genome sequence of Vibrio vulnificus CMCP6.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Chaperone subunit of a proteasome-like degradation  
 CC complex (By similarity).  
 CC -1- SUBUNIT: A double ring-shaped homohexameric of hslu is capped on  
 CC each side by a ring-shaped hslu homohexameric (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: Belongs to the clpX chaperone family. Hslu subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE016801; AA009806.1; -.  
 DR HAMAP: MF 00249; -; 1.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003959; AAA\_ATPase\_central.  
 DR InterPro: IPR004491; Hsp\_HsluV.  
 DR Pfam: PF00004; AAA; 1.  
 DR SMART: SM00382; AAA; 1.  
 DR TIGRFAMs: TIGR00390; hslu; 1.  
 DR KW Chaperone; ATP-binding; Complete proteome.  
 FT NP\_BIND 57 64 ATP (POTENTIAL).  
 SQ SEQUENCE 443 AA; 49993 MW; 634F55B3AC85F23D CRC64;  
 Query Match 48.1%; Score 49.5; DB 1; Length 443;  
 Best Local Similarity 41.7%; Pred. No. 3.9;  
 Matches 10; Conservative 6; Mismatches 3; Indels 5; Gaps 1;  
 Qy 1 ORLKYAQBELSNE-----VLPPPP 19  
 Db 117 EKVFRAEELAEERVLDAALLPPPP 140  
 RESULT 15  
 RPB2\_HUMAN  
 ID RPB2\_HUMAN STANDARD; PRT; 1174 AA.  
 AC P30876;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE DNA-directed RNA polymerase II 140 kDa polypeptide (EC 2.7.7.6)  
 DE (RNA polymerase II subunit 2) (RPS2).  
 GN POLR2B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92389336; PubMed=1518060;  
 RA Acker J., Wintzerith M., Vigneron M., Keding C.;  
 RT "Primary structure of the second largest subunit of human RNA  
 RT polymerase II (or B).";  
 RL J. Mol. Biol. 226:1295-1299(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

DR Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RA "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RN SEQUENCE OF 699-1174 FROM N.A.  
RN TISSUE=Brain;  
RC MEDLINE=97264341; PubMed=9110174;  
RX Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,  
RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;  
RT "Large-scale concatenation cDNA sequencing."  
RL Genome Res. 7:353-358(1997).  
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
CC of DNA into RNA using the four ribonucleoside triphosphates as  
CC substrates.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC [RNA] (N).  
CC -!- SUBUNIT: RNA polymerase II consists of 12 different subunits. This  
CC subunit is the second largest component of RNA polymerase II.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases are  
CC found in eukaryotic nuclei: polymerase I for the ribosomal RNA  
CC precursor, polymerase II for the mRNA precursor, and polymerase  
CC III for 5S and tRNA genes.  
CC -!- SIMILARITY: Belongs to the RNA polymerase beta chain family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X63563; CAA45124.1; -;  
DR EMBL; BC023503; AAR23503.2; -;  
DR EMBL; AF055028; AAC09367.1; -;  
DR PIR; S28976; S28976.  
DR Genew; HGNC:9188; POLR2B.  
DR MIM; 180661; -;  
DR GO; GO:0005665; C:DNA-directed RNA polymerase II, core complex; TAS.  
DR GO; GO:0003677; F:DNA binding; TAS.  
DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.  
DR InterPro; IPR007121; RNA\_pol\_B.  
DR InterPro; IPR007644; RNA\_pol\_Rpb2\_1.  
DR InterPro; IPR007642; RNA\_pol\_Rpb2\_2.  
DR InterPro; IPR007645; RNA\_pol\_Rpb2\_3.  
DR InterPro; IPR007646; RNA\_pol\_Rpb2\_4.  
DR InterPro; IPR007647; RNA\_pol\_Rpb2\_5.  
DR InterPro; IPR007120; RNA\_pol\_Rpb2\_6.  
DR InterPro; IPR007641; RNA\_pol\_Rpb2\_7.  
DR Pfam; PF04563; RNA\_pol\_Rpb2\_1; 1.  
DR Pfam; PF04561; RNA\_pol\_Rpb2\_2; 1.  
DR Pfam; PF04565; RNA\_pol\_Rpb2\_3; 1.  
DR Pfam; PF04566; RNA\_pol\_Rpb2\_4; 1.  
DR Pfam; PF04567; RNA\_pol\_Rpb2\_5; 1.  
DR Pfam; PF00562; RNA\_pol\_Rpb2\_6; 1.  
DR Pfam; PF04560; RNA\_pol\_Rpb2\_7; 1.

DR PROSITE; PS01166; RNA\_POL\_BETA; 1.  
KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;  
FT Zinc-finger; Nuclear protein.  
FT ZN FING 1119 1140 C4-TYPE (POTENTIAL).  
SQ SEQUENCE 1174 AA; 133896 MW; 32BED7F95E4DE10 CRC64; -  
Query Match 47.6%; Score 49; DB 1; Length 1174;  
Best Local Similarity 56.2%; Pred. No. 14;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVL 16  
DB 334 KRIKYAKEVLQKEMLP 349  
RESULT 16  
NAFL\_HUMAN  
ID NAFL\_HUMAN STANDARD; PRT; 636 AA.  
AC Q15025; O76008; Q96EL9; Q99833; Q9HLJ3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Nef-associated factor 1 (Naf1) (HIV-1 Nef interacting protein)  
DE (Viron-associated nuclear shuttling protein) (VAN) (hVAN) (TNFAIP3  
DE interacting protein 1).  
GN TNF1 OR NAF1 OR KIAA0113.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC TISSUE=Peripheral blood;  
RX MEDLINE=99120485; PubMed=9923610;  
RA Fukushi M., Dixon J., Kimura T., Tsurutani N., Dixon M.J.,  
RA Yamamoto N.;  
RT "Identification and cloning of a novel cellular protein Naf1, Nef-  
RT associated factor 1, that increases cell surface CD4 expression."  
RL FEBS Lett. 442:83-88(1999).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Leukocyte;  
RX MEDLINE=20541981; PubMed=11090181;  
RA Gupta K., Ott D., Hope T.J., Siliciano R.F., Boeke J.D.;  
RT "A human nuclear shuttling protein that interacts with human  
RT immunodeficiency virus type 1 matrix is packaged into virions."  
RL J. Virol. 74:11811-11824(2000).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RL human and mouse cDNA sequences."  
RN [4]  
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RP SEQUENCE OF 136-636 FROM N.A. (ISOFORM 2).

CC TISSUE=Craniofacial;  
RX MEDLINE=96276047; PubMed=8681136;  
RA Loftus S.K., Dixon J., Kopriwnikar K., Dixon M.J., Wasmuth J.J.;  
RT "Transcriptional map of the Treacher Collins candidate gene region.";  
RL Genome Res. 6:26-34(1996).  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Chara O., Nagase T., Kikuno R., Nomura N.;  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE OF 341-636 FROM N.A. (ISOFORM 1).  
RC TISSUE=Bone marrow;  
RX MEDLINE=95308325; PubMed=7798527;  
RA Nagase T., Miyajima N., Tanaka A., Sazuka T., Seki N., Sato S.,  
RA Tabata S., Ishikawa K.-I., Kawarabayashi Y., Kotani H., Nomura N.;  
RT "Prediction of the coding sequences of unidentified human genes. III.  
RT The coding sequences of 40 new genes (K1AA0081-K1AA0120) deduced by  
RT analysis of cDNA clones from human cell line KG-1.";  
RL DNA Res. 2:37-43(1995).  
RN [7]  
RP SEQUENCE OF 94-412 FROM N.A.  
RA Fukushi M., Kimura T., Yamamoto N.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Interacts with zinc finger protein A20/TNFAIP3 and  
CC inhibits TNF-induced NF-kappa-B-dependent gene expression by  
CC interfering with an RIP- or TRAF2-mediated transactivation signal  
CC (By similarity). Increases cell surface CD4(T4) antigen  
CC expression. Interacts with HIV-1 matrix protein and is packaged  
CC into virions and overexpression can inhibit viral replication. May  
CC regulate matrix nuclear localization, both nuclear import of PIC  
CC (preintegration complex) and export of GAG polyprotein and viral  
CC genomic RNA during virion production.  
CC -1- SUBUNIT: Interacts with TNFAIP3 (By similarity). Interacts with  
CC HIV-1 matrix protein.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Shuttles between the nucleus  
CC and cytoplasm in a CRM1-dependent manner.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing; Named isoforms=2;  
CC Name=1; Synonyms=Alpha;  
CC IsoId=Q15025-1; Sequences=Displayed;  
CC Name=2; Synonyms=Beta;  
CC IsoId=Q15025-2; Sequences=VSP 003913;  
CC Note=No experimental confirmation available;  
CC -1- TISSUE SPECIFICITY: Ubiquitous. Strongly expressed in peripheral  
CC blood lymphocytes, spleen and skeletal muscle, and is weakly  
CC expressed in the brain.  
CC -1- CAUTION: Ref.7 sequence differs from that shown due to frameshifts  
CC in positions 152 and 154.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ011895; CAA09855.1; -;  
DR EMBL; AJ011896; CAA09856.1; -;  
DR EMBL; AY012155; AAG42154.1; -;  
DR EMBL; BC012133; AAH12133.1; -;  
DR EMBL; BC014008; AAH14008.1; -;  
DR EMBL; U39403; AAC99999.1; -;  
DR EMBL; D30755; BAA06416.2; -;  
DR EMBL; U03844; AAB41438.1; ALT\_FRAME.  
DR Genbank; HGNC:16903; TNIP1.  
DR MIM; 607714; -;  
DR GO; GO:0005622; C:intracellular; TAS.  
DR GO; GO:0005315; F:protein binding; TAS.  
DR GO; GO:0009101; P:glycoprotein biosynthesis; IDA.  
DR GO; GO:0045071; P:negative regulation of viral genome replication; TAS.  
DR GO; GO:0045071; P:negative regulation of viral genome replication; TAS.  
KW Coiled coil; Nuclear protein; Alternative splicing.  
FT DOMAIN 20 73 COILED COIL (POTENTIAL).

FT DOMAIN 196 258 COILED COIL (POTENTIAL).  
FT DOMAIN 294 535 COILED COIL (POTENTIAL).  
FT DOMAIN 94 412 INTERACTS WITH NEF.  
FT DOMAIN 524 530 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 539 636 PRO-RICH.  
FT VARSPLIC 627 636 SPKNDREGPQ -> PADLRPRN (in isoform 2).  
FT CONFLICT 148 148 G -> D (IN REF. 3; AAH12133).  
FT CONFLICT 299 299 A -> P (IN REF. 2).  
SQ SEQUENCE 636 AA; 71864 MW; D81B96BEAD50D871 CRC64;  
Query Match 46.6%; Score 48; DB 1; Length 636;  
Best Local Similarity 57.9%; Pred.No. 9.9;  
Matches 11; Conservative 2; Mismatches 4; Indels 2; Gaps 1;  
Oy 2 RLKYAQEEL--SNEVLPPP 18  
||: ||| ||: |||  
Db 58 RLKQAEELVKNELPPP 76  
RESULT 17  
RFB2\_DROME  
ID RPB2\_DROME STANDARD; PRT; 1176 AA.  
AC P08266; Q04155; Q95027; Q9VFM7;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE DNA-directed RNA polymerase II 140 kDa polypeptide (EC 2.7.7.6)  
DE (RNA polymerase II subunit 2).  
GN RPII140 OR Q93180.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=88011299; PubMed=3116266;  
RA Falkenburg D., Dworniczak B., Faust D.M., Bautz E.K.F.;  
RT "RNA polymerase II of Drosophila. Relation of its 140,000 Mr subunit  
RT to the beta subunit of Escherichia coli RNA polymerase.";  
RL J. Mol. Biol. 195:929-937(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Calle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laeko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,



DR PIR; S19503; LLBY.  
 DR PDB; 1HQZ; 21-DEC-01.  
 DR PDB; 1JQ8; 01-MAR-02.  
 DR GermOnline; 138985; -.  
 DR SGM; S0000684; ABP1.  
 DR GO; GO:0005938; C:cell cortex; IDA.  
 DR GO; GO:0005737; Cytoplasm; IDA.  
 DR GO; GO:0005515; F:protein binding; IDA.  
 DR GO; GO:0000147; P:actin cortical patch assembly; IMP.  
 DR GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. .; IMP.  
 DR InterPro; IPR002108; Actbind cofln.  
 DR InterPro; IPR000108; Neu\_cyt\_fact\_2.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00241; Cofilin\_ADF; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR00499; P67PHOX.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR ProDom; PD000066; SH3; 1.  
 DR SMART; SM00102; ADF; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; P05002; SH3; 1.  
 KW Cytoskeleton; Actin-binding; SH3 domain; Repeat; 3D-structure.  
 FT DOMAIN 86 97 ACTIN-BINDING (POTENTIAL).  
 FT DOMAIN 532 592 SH3.  
 FT FT DOMAIN 200 575 3 X 10 AA APPROXIMATE REPEATS.  
 FT REPEAT 200 209 1.  
 FT REPEAT 436 445 2.  
 FT REPEAT 566 575 3.  
 FT CONFLICT 58 58 L -> S (IN REF. 1).  
 FT CONFLICT 312 312 K -> I (IN REF. 1).  
 SQ SEQUENCE 592 AA; 65576 MW; 39523510704D94AA CRC64;  
  
 Query Match 45.6%; Score 47; DB 1; Length 592;  
 Best Local Similarity 50.0%; Pred. No. 13;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 7 QEELSNEVLPPPK 20  
 :||:|:||||:  
 Db 369 EEEMENKFAPPPKK 382  
  
 RESULT 20  
 UN89 CAEL STANDARD; PRT; 6632 AA.  
 ID UN89 CAEL 001761; Q17362;  
 AC 001761; Q17362;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).  
 GN UNC-89 OR C09D1.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Perodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=96180278; PubMed=9603916;  
 RA Benian G.M., Tinley T.B., Tang X., Borodovsky M.;  
 RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line  
 ET assembly, encodes a giant modular protein composed of Ig and signal  
 ET transduction domains".  
 RL J. Cell Biol. 132:835-848 (1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Du Z., Le T.T., Wilson R.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP REVISIONS.  
 RA Waterston R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Structural component of the muscle M-line. Myofibrament

lattice assembly begins with positional cues laid down in the basement membrane and muscle cell membranes. UNC-89 responds to these signals, localizes, and then participates in assembling an M-line.

-!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.

-!- SIMILARITY: Contains 1 DBL-homology (DB) domain.

-!- SIMILARITY: Contains 1 fibronectin type III domain.

-!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.

-!- SIMILARITY: Contains 1 PH domain.

-!- SIMILARITY: Contains 5 RCSD domains.

-!- SIMILARITY: Contains 1 SH3 domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; U33058; AAB0542.1; -.  
EMBL; AF003131; AAB54132.2; -.  
PDB; 1FHO; 20-DEC-00.  
WormPep; C09D1.1; CE30426.  
InterPro; IPR008957; FN\_III-like.  
InterPro; IPR003961; FN\_III.  
InterPro; IPR007110; IG-like.  
InterPro; IPR003598; IG\_C2.  
InterPro; IPR003006; IG\_MHC.  
InterPro; IPR001849; PH.  
InterPro; IPR007850; RCSD.  
InterPro; IPR000219; RhGEF.  
InterPro; IPR001452; SH3.  
Pfam; PF00041; fn3; 1.  
Pfam; PF00047; ig; 47.  
Pfam; PF00169; PH; 1.  
Pfam; PF05177; RCSD; 5.  
Pfam; PF00621; RhGEF; 1.  
Pfam; PF00018; SH3; 1.  
SMART; SM00408; IGc2; 23.  
SMART; SM00325; RhGEF; 1.  
SMART; SM00326; SH3; 1.  
PROSITE; PS50010; DH\_2; 1.  
PROSITE; PS50835; IG\_LIKE; 49.  
PROSITE; PS50003; PH\_DOMAIN; 1.  
PROSITE; PS50002; SH3; 1.  
Muscle protein; immunoglobulin domain; Repeat; SH3 domain; 3D-structure.  
DOMAIN 63 127 SH3.  
DOMAIN 152 330 DH.  
DOMAIN 342 498 PH.  
DOMAIN 547 633 IG-LIKE C2-TYPE 1.  
DOMAIN 648 736 IG-LIKE C2-TYPE 2.  
DOMAIN 748 838 IG-LIKE C2-TYPE 3.  
DOMAIN 946 1033 IG-LIKE C2-TYPE 4.  
DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.  
DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.  
DOMAIN 1272 1315 THR-RICH.  
DOMAIN 1375 1475 RCSD 1.  
DOMAIN 1479 1585 RCSD 2.  
DOMAIN 1597 1685 RCSD 3.  
DOMAIN 1700 1799 RCSD 4.  
DOMAIN 1800 1860 RCSD 5.  
DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.  
DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.  
DOMAIN 2171 2281 IG-LIKE C2-TYPE 9.  
DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.  
DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.  
DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.  
DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.  
DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.  
DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.  
DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.

FT	DOMAIN	2994	3081	IG-LIKE C2-TYPE 17.
FT	DOMAIN	3087	3183	IG-LIKE C2-TYPE 18.
FT	DOMAIN	3189	3280	IG-LIKE C2-TYPE 19.
FT	DOMAIN	3286	3376	IG-LIKE C2-TYPE 20.
FT	DOMAIN	3384	3469	IG-LIKE C2-TYPE 21.
FT	DOMAIN	3482	3572	IG-LIKE C2-TYPE 22.
FT	DOMAIN	3580	3667	IG-LIKE C2-TYPE 23.
FT	DOMAIN	3686	3777	IG-LIKE C2-TYPE 24.
FT	DOMAIN	3817	3908	IG-LIKE C2-TYPE 25.
FT	DOMAIN	3920	4009	IG-LIKE C2-TYPE 26.
FT	DOMAIN	4018	4106	IG-LIKE C2-TYPE 27.
FT	DOMAIN	4109	4201	IG-LIKE C2-TYPE 28.
FT	DOMAIN	4212	4297	IG-LIKE C2-TYPE 29.
FT	DOMAIN	4302	4387	IG-LIKE C2-TYPE 30.
FT	DOMAIN	4400	4485	IG-LIKE C2-TYPE 31.
FT	DOMAIN	4489	4580	IG-LIKE C2-TYPE 32.
FT	DOMAIN	4588	4678	IG-LIKE C2-TYPE 33.
FT	DOMAIN	4681	4771	IG-LIKE C2-TYPE 34.
FT	DOMAIN	4873	4961	IG-LIKE C2-TYPE 35.
FT	DOMAIN	4965	5057	IG-LIKE C2-TYPE 36.
FT	DOMAIN	5067	5160	IG-LIKE C2-TYPE 37.
FT	DOMAIN	5171	5260	IG-LIKE C2-TYPE 38.
FT	DOMAIN	5277	5366	IG-LIKE C2-TYPE 39.
FT	DOMAIN	5383	5472	IG-LIKE C2-TYPE 40.
FT	DOMAIN	5487	5578	IG-LIKE C2-TYPE 41.
FT	DOMAIN	5595	5685	IG-LIKE C2-TYPE 42.
FT	DOMAIN	5701	5790	IG-LIKE C2-TYPE 43.
FT	DOMAIN	5815	5904	IG-LIKE C2-TYPE 44.
FT	DOMAIN	5925	6014	IG-LIKE C2-TYPE 45.
FT	DOMAIN	6038	6130	IG-LIKE C2-TYPE 46.
FT	DOMAIN	6150	6239	IG-LIKE C2-TYPE 47.
FT	DOMAIN	6275	6368	FIBRONECTIN TYPE-III.
FT	DOMAIN	6413	6502	IG-LIKE C2-TYPE 48.
FT	DOMAIN	6507	6596	IG-LIKE C2-TYPE 49.
FT	DISULFID	568	621	POTENTIAL.
FT	DISULFID	2908	2975	POTENTIAL.
FT	DISULFID	3015	3085	POTENTIAL.
FT	DISULFID	3707	3759	POTENTIAL.
FT	DISULFID	3826	3890	POTENTIAL.
FT	DISULFID	5092	5157	POTENTIAL.
FT	DISULFID	5298	5350	POTENTIAL.
FT	DISULFID	5508	5560	POTENTIAL.
FT	DISULFID	5616	5669	POTENTIAL.
FT	DISULFID	5722	5764	POTENTIAL.
FT	DISULFID	5836	5901	POTENTIAL.
FT	DISULFID	5946	5998	POTENTIAL.
FT	DISULFID	6036	6171	POTENTIAL.
FT	DISULFID	6421	6486	POTENTIAL.
FT	CONFLICT	2137	2137	A -> P (IN REF. 1).
FT	CONFLICT	2245	2247	AKA -> PKP (IN REF. 1).
FT	CONFLICT	2258	2258	A -> P (IN REF. 1).
FT	CONFLICT	2284	2284	E -> G (IN REF. 1).
FT	CONFLICT	2297	2297	M -> I (IN REF. 1).
FT	CONFLICT	3531	3531	A -> G (IN REF. 1).
FT	CONFLICT	3884	3888	DAGEY -> RRRRI (IN REF. 1).
FT	CONFLICT	3929	3929	A -> V (IN REF. 1).
FT	CONFLICT	5134	5134	A -> P (IN REF. 1).
FT	CONFLICT	5145	5145	T -> S (IN REF. 1).
FT	CONFLICT	5185	5185	G -> A (IN REF. 1).
FT	CONFLICT	5199	5199	K -> N (IN REF. 1).
FT	CONFLICT	5202	5202	L -> F (IN REF. 1).
FT	CONFLICT	5213	5213	P -> L (IN REF. 1).
FT	CONFLICT	6178	6178	A -> G (IN REF. 1).
FT	CONFLICT	6268	6268	K -> E (IN REF. 1).
FT	CONFLICT	6532	6532	AA; 731665 MW; 262D3EDD62960B89 CRC64;
FT	SEQUENCE	6532	6532	AA; 731665 MW; 262D3EDD62960B89 CRC64;

Query Match 45.6%; Score 47; DB 1; Length 6632;  
Best Local Similarity 45.0%; Pred. No. 1.9e+02;  
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPPRK 20  
Db 1353 RRVSAEEELPKVIDSDRK 1372

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

CC EMBL; S73885; AAB32235.1; --  
CC EMBL; AC004653; AAC17116.1; --  
CC EMBL; BC010576; AAH10576.1; --  
CC EMBL; X57435; CAA40683.1; --  
CC PIR; I56893; I56893.  
CC HSSP; P22415; 1AN4.  
CC TRANSFAC; T00036; --  
CC Genew; HGNC:11745; TFAP4.  
CC MIM; 600743; --  
CC GO; GO:0003705; F:RNA polymerase II transcription factor acti. .; TAS.  
CC GO; GO:0003713; P:transcription co-activator activity; TAS.  
CC GO; GO:0007275; P:development; TAS.  
CC GO; GO:0006357; P:regulation of transcription from Pol II pro. .; TAS.  
CC InterPro; IPR001092; HLH\_basic.  
CC SMART; SMO010; HLH; 1.  
CC PROSITE; PS00888; HLH; 1.  
CC Transcription regulation; DNA-binding; Activator; Nuclear protein.  
KW DNA BIND 48 60 BASIC DOMAIN.  
FT DOMAIN 61 100 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
FT DOMAIN 99 120 LEUCINE-ZIPPER 1.  
FT DOMAIN 151 179 LEUCINE-ZIPPER 2.  
FT DOMAIN 193 222 GLN-RICH.  
FT DOMAIN 225 244 PRO-RICH.  
FT CONFLICT 218 218 Q -> H (IN REF. 2). CRC64;  
SQ SEQUENCE 338 AA; 38725 MW; 540C00856596B83 CRC64;

Query Match 44.7%; Score 46; DB 1; Length 338;  
Best Local Similarity 38.9%; Pred. No. 10;  
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 ORLYAQEELSNEVLPPP 18  
::: |||: |||:  
Db 210 EKLEREQQLRTQLPPP 227

RESULT 22  
11SB CUCMA STANDARD; PRT; 480 AA.  
ID 11SB CUCMA STANDARD; PRT; 480 AA.  
AC P1374;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 11S globulin beta subunit precursor.  
OS Cucurbita maxima (Pumpkin) (Winter squash).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucurbita.  
OX NCBI\_TaxID=3661;  
RN [1] \_SEQUENCE FROM N.A.  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Kurokawa Anakuri Nankin;  
RX MEDLINE=8166744; PubMed=2450746;  
RA Hayashi M., Mori H., Nishimura M., Akazawa T., Hara-Nishimura I.;  
RT "Nucleotide sequence of cloned cDNA coding for pumpkin 11-S globulin  
RT beta subunit."  
RL Eur. J. Biochem. 172:627-632(1988).  
RN [2]  
RP SEQUENCE OF 22-30 AND 297-302.  
RA Ohmiya M., Hara I., Mastubara H.;  
RT "Pumpkin (Cucurbita sp.) seed globulin IV. Terminal sequences of the  
RT acidic and basic peptide chains and identification of a pyroglutamy  
RT peptide chain."  
RL Plant Cell Physiol. 21:157-167(1980).  
CC -!- FUNCTION: This is a seed storage protein.  
CC -!- SUBUNIT: Hexamer; each subunit is composed of an acidic and a  
CC basic chain derived from a single precursor and linked by a  
CC disulfide bond.  
CC -!- SIMILARITY: Belongs to the 11S seed storage protein (globulins)

RESULT 21  
TAP4 HUMAN STANDARD; PRT; 338 AA.  
ID TAP4 HUMAN STANDARD; PRT; 338 AA.  
AC Q01664; O60409;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1995 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Transcription factor AP-4 (Activating enhancer-binding protein 4).  
GN TFAP4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95018629; PubMed=7933101;  
RA Ou S.H., Garcia-Martinez L.F., Paulsen E.J., Gaynor R.B.;  
RT "Role of flanking E box motifs in human immunodeficiency virus type 1  
RT TATA element function."  
RL J. Virol. 68:7188-7199(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,  
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,  
RA Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,  
RA Ueng S., Tatum O., Campbell C., Fawcett J., Maltbie M., Misra M.,  
RA Deaven L.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Brain;  
MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Tohiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).  
RN [4]  
RP SEQUENCE OF 18-338 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=91065520; PubMed=2123466;  
RA Hu Y.-F., Luescher B., Admon A., Mermoud N., Tian R.;  
RT "Transcription factor AP-4 contains multiple dimerization domains  
RT that regulate dimer specificity."  
RL Genes Dev. 4:1741-1752(1990).  
CC -!- FUNCTION: Transcription factor that activates both viral and  
CC cellular genes by binding to the symmetrical DNA sequence  
CC 5'-CAGTGTG-3'.  
CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another  
CC bHLH protein, Homodimer.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M36407; AAA33110.1; --

InterPro; IPR006045; Cupin.

InterPro; IPR007113; Cupin sup.

InterPro; IPR006044; SeedStore\_11s.

Pfam; PF00190; Cupin; 2.

PRINTS; PR00439; 11SGLOBULIN.

PROSITE; PS00305; 11S\_SEED\_STORAGE; 1.

Seed storage protein; Signal; Pyrrolidone carboxylic acid.

FT SIGNAL 1 21

FT CHAIN 22 480 11S GLOBULIN BETA SUBUNIT.

FT CHAIN 22 296 GAMMA CHAIN (ACIDIC).

FT CHAIN 297 480 DELTA CHAIN (BASIC).

FT MOD RES 22 22 PYRROLIDONE CARBOXYLIC ACID.

FT DISULFID 124 303 INTERCHAIN (GAMMA-DELTA) (POTENTIAL).

FT CONFLICT 27 27 S -> E (IN REF. 2).

FT CONFLICT 30 30 E -> S (IN REF. 2).

SEQUENCE 480 AA; 54625 MW; BCD8A83DD1AED93C CRC64;

Query Match 44.7%; Score 46; DB 1; Length 480;

Best Local Similarity 57.9%; Pred. No. 15;

Matches 11; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 1 QRLKYAQEELSNEVLPPPP 19

Db 458 QRLKYQEQEM--RVLSPCR 474

RESULT 23

HEXA BLADI

ID HEXA BLADI STANDARD; PRT; 733 AA.

AC Q17127;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Hexamerin precursor.

OS Blaberus discoidalis (Tropical cockroach).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blaberoidea;

OC Blaberidae; Blaberus.

OC NCBI\_TaxID=6981;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pat body;

RA Jamroz R.C., Beintema J.J., Stam W.T., Bradfield J.Y.;

RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Larval storage protein (LSP) which may serve as a store of amino acids for synthesis of adult proteins (By similarity).

CC -!- SUBUNIT: Homohexamer (Potential).

CC -!- SUBCELLULAR LOCATION: Extracellular (By similarity).

CC -!- SIMILARITY: TO ARYLPHORINS AND TO ARTHROPOD HEMOCYANINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U01328; AAA74579.1; --

DR HSSP; P04253; 10XY.

InterPro; IPR008922; Di-copper centre.

InterPro; IPR00896; Hemocyanin.

DR InterPro; IPR005203; hemocyanin C.

DR InterPro; IPR005204; hemocyanin N.

DR InterPro; IPR007110; Ig-like.

DR Pfam; PF00372; hemocyanin; 1.

DR Pfam; PF03723; hemocyanin C; 1.

DR Pfam; PF03722; hemocyanin N; 1.

DR PRINTS; PR00187; HAEMOCYANIN.

DR PROSITE; PS00209; HEMOCYANIN 1; FALSE NEG.

DR PROSITE; PS00210; HEMOCYANIN 2; FALSE NEG.

KW Signal; Storage protein; Glycoprotein.

FT SIGNAL 1 17 POTENTIAL.

FT CHAIN 18 733 HEXAMERIN.

FT CARBOHYD 199 199 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 431 431 N-LINKED (GLCNAC. .) (POTENTIAL).

SQ SEQUENCE 733 AA; 87813 MW; 083DF739DD665729 CRC64;

Query Match 43.7%; Score 45; DB 1; Length 733;

Best Local Similarity 50.0%; Pred. No. 34;

Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 5 YAQEELSNEVLPPPP 18

Db 148 FQEDLSYILPPPP 161

RESULT 24

RPB2 ARATH

ID RPB2 ARATH STANDARD; PRT; 1188 AA.

AC P38420; Q9SVS6;

DT 01-OCT-1994 (Rel. 30, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE DNA-directed RNA polymerase II 135 kDa polypeptide (EC 2.7.7.6)

DE (RNA polymerase II subunit 2).

GN RPB135 OR RPB2 OR RPI40 OR AT4G21710 OR F17L22.170.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OC NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G., Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N., Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M., Mueller M., Weichselgartner M., de Simone V., Obermaier B., Maché R., Watson M., Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T., Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I., Vos P., Hobeisel J., Zimmermann W., Wedler H., Ridley P., Langham S.-A., McCullagh B., Bilham L., Robben J., Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F., Braeken M., Weijens I., Voet M., Bastiaens I., Aert R., Defoor E., Weitzenecker T., Bothe G., Rampsperger U., Hilbert H., Braun M., Holzer E., Brandt A., Peters S., van Staveren M., Dirks W., Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetter P., Bernsner S., Hempel S., Feldpausch M., Lambirth S., Van den Daele H., De Keyser A., Buyschaert C., Gielen J., Cronin A., Quail M.A., Bray-Allen R., Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen R., Clark L., Doggett J., Hall S., Kay M., Lennard N., McIlroy K., Mayes R., Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S., Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H., Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,



FT ZN\_PING 1124 1145 C4-TYPE (POTENTIAL).  
FT VARIANT 787 787 I -> N.  
FT CONFLICT 354 355 PH -> LY (IN REF. 1).  
SQ SEQUENCE 1188 AA; 135018 MW; C304E43515C2C364 CRC64;  
Query Match 43.7%; Score 45; DB 1; Length 1188;  
Best Local Similarity 50.0%; Pred. No. 59; Indels 0;  
Matches 8; Conservative 5; Mismatches 3; Gaps 0;  
QY 1 ORLKYAQEELSNEVLP 16  
:|||||:|:|  
DB 339 KRIKYARDILQKEMLP 354  
:|||||:|:|  
RESULT 25  
PMG3 YEAST  
ID PMG3 YEAST STANDARD; PRT; 303 AA.  
AC Q12326;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Phosphoglycerate mutase 3 (EC 5.4.2.1) (phosphoglyceromutase 3) (PGAM  
DE 3) (MPEGM 3) (BPG-dependent PGAM 3).  
GN GPM3 OR YOL056W OR O1236.  
OS Saccharomycetes cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / FY73;  
RX MEDLINE=963181248; PubMed=8789261;  
RA Manhaupt G., Vetter I., Schwarzlose C., Mittel S., Feldmann H.;  
RT "Analysis of a 26 kb region on the left arm of yeast chromosome XV.";  
RL Yeast 12:67-76(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Anseorge W., Benes V., Rechmann S., Schwager C., Teodoru C., Voss H.,  
RA Wiemann S.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=98205882; PubMed=9544241;  
RA Heinsch J.J., Mueller S., Schluter E., Jacoby J., Rodicio R.;  
RT "Investigation of two yeast genes encoding putative isoenzymes of  
RT phosphoglycerate mutase."  
RL Yeast 14:203-213(1998).  
CC -!- FUNCTION: Could be non-functional.  
CC -!- CATALYTIC ACTIVITY: 2-phospho-D-glycerate + 2,3-diphosphoglycerate  
CC = 3-phospho-D-glycerate + 2,3-diphosphoglycerate.  
CC -!- PATHWAY: Glycolysis.  
CC -!- SIMILARITY: Belongs to the phosphoglycerate mutase family. BPG-  
CC dependent PGAM subfamily.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X91067; CAA62530.1; -.  
DR EMBL; Z74798; CAA99064.1; -.  
DR PIR; S61723; S61723.  
DR HSSP; P00950; 5PGM.  
DR GERMOnline; 143479; -.  
DR SGD; S0005417; GPM3.  
DR GO; GO:0004619; F:phosphoglycerate mutase activity; IMP.  
DR InterPro; IPR001345; PG/BPGM mutase.  
DR InterPro; IPR005952; Phosphogly\_mut1.  
DR Pfam; PF00300; PGAM; 1.  
DR TIGRFAme; TIGR01258; pgm\_1; 1.

RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,  
RA Neumann S., Argirion A., Vitale D., Liguori R., Piravandi E.,  
RA Massenon O., Quigley F., Clabaud G., Muendlein A., Feiber R.,  
RA Schnabel S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
RA Chedfor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,  
RA Perez-Perez A., Fumelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
RA Heijnen L., Schwarz S., Scholler P., Heber S., Frances P., Bielke C.,  
RA Frishman D., Haase D., Lencke K., Mewes H.-W., Stocker S.,  
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
RA Parnell L., Dedhia N., Guoj L., Schutz K., Huang E., Spiegel L.,  
RA Sekhon M., Murray J., Shest P., Cordes M., Abu-Threiden J.,  
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
RA Kramer J., Spiech J., Ryan E., Andrews S., Geisel C., Layman D.W.,  
RA Nelson J., Ali J., Bernhoff A., Jones K., Drone K., Cotton M., Joshi C.,  
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
RA Chen E., Maria M.A., Martensen R., McCombie W.R.;  
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
RT thaliana."  
RL Nature 402:769-777(1999).  
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
CC of DNA into RNA using the four ribonucleoside triphosphates as  
CC substrates.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC {RNA} (N).  
CC -!- SUBUNIT: Each class of RNA polymerase is assembled from 9 to 14  
CC different polypeptides. This subunit is the second largest  
CC component of RNA polymerase II.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases are  
CC found in eukaryotic nuclei: polymerase I for the ribosomal RNA  
CC precursor, polymerase II for the mRNA precursor, and polymerase  
CC III for SS and tRNA genes.  
CC -!- SIMILARITY: Belongs to the RNA polymerase beta chain family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z19120; CAA79527.1; -.  
DR EMBL; Z19121; CAA79528.1; -.  
DR EMBL; AL035527; CAB36815.1; -.  
DR EMBL; AL161555; CAB81278.1; -.  
DR PIR; T05846; T05846.  
DR InterPro; IPR007121; RNA\_pol\_B.  
DR InterPro; IPR007644; RNA\_pol\_Rpb2\_1.  
DR InterPro; IPR007642; RNA\_pol\_Rpb2\_2.  
DR InterPro; IPR007645; RNA\_pol\_Rpb2\_3.  
DR InterPro; IPR007646; RNA\_pol\_Rpb2\_4.  
DR InterPro; IPR007647; RNA\_pol\_Rpb2\_5.  
DR InterPro; IPR007120; RNA\_pol\_Rpb2\_6.  
DR InterPro; IPR007641; RNA\_pol\_Rpb2\_7.  
DR Pfam; PF04563; RNA\_pol\_Rpb2\_1; 1.  
DR Pfam; PF04561; RNA\_pol\_Rpb2\_2; 1.  
DR Pfam; PF04565; RNA\_pol\_Rpb2\_3; 1.  
DR Pfam; PF04566; RNA\_pol\_Rpb2\_4; 1.  
DR Pfam; PF04567; RNA\_pol\_Rpb2\_5; 1.  
DR Pfam; PF00562; RNA\_pol\_Rpb2\_6; 1.  
DR Pfam; PF04560; RNA\_pol\_Rpb2\_7; 1.  
DR PROSITE; PS01166; RNA\_POL\_BETA; 1.  
KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;  
KW Zinc-finger; Nuclear protein.  
FT DOMAIN 14 20 ASP/GLU-RICH (ACIDIC).







RP SEQUENCE FROM N.A.  
RC STRAIN=Golden retriever;  
RA Carville K.S., Mann C.J., Schatzberg S.J., Wilton S.D.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: May play a role in anchoring the cytoskeleton to the  
CC plasma membrane.  
CC -!- SUBUNIT: Interacts with the syntrophins SNTA1, SNTB1, SNTB2, SNTG1  
CC and SNTG2 (By similarity).  
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY  
CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,  
CC ABP-120, ABP-180, OR BETA-FODRIN).  
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
CC -!- SIMILARITY: Contains 22 spectrin repeats.  
CC -!- SIMILARITY: Contains 1 WW domain.  
CC -!- SIMILARITY: Contains 1 ZZ-type zinc finger.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF070485; AAC83646.1; -.  
DR HSSP; P46939; IQAG.  
DR InterPro; IPR001589; Actbind\_actnin.  
DR InterPro; IPR001715; Calponin-like.  
DR InterPro; IPR002017; Spectrin.  
DR InterPro; IPR001202; WW Rep5\_WWP.  
DR InterPro; IPR000433; ZnF\_ZZ.  
DR Pfam; PF00307; CH; 2.  
DR Pfam; PF00435; spectrin; 21.  
DR Pfam; PF00397; WW; 1.  
DR Pfam; PF00569; ZZ; 1.  
DR SMART; SM00033; CH; 2.  
DR SMART; SM00150; SPEC; 21.  
DR SMART; SM00456; WW; 1.  
DR SMART; SM00291; ZnF\_ZZ; 1.  
DR PROSITE; PS00019; ACTININ\_1; 1.  
DR PROSITE; PS00020; ACTININ\_2; 1.  
DR PROSITE; PS50021; CH; 2.  
DR PROSITE; PS01159; WW\_DOMAIN\_1; 1.  
DR PROSITE; PS00020; WW\_DOMAIN\_2; 1.  
DR PROSITE; PS01357; ZF\_ZZ\_1; 1.  
DR PROSITE; PS0135; ZF\_ZZ\_2; 1.  
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;  
KW Repeat; Zinc-finger.  
FT DOMAIN 15 119 CH 1.  
FT DOMAIN 134 237 CH 2.  
FT REPEAT 340 448 SPECTRIN 1.  
FT REPEAT 449 557 SPECTRIN 2.  
FT REPEAT 560 668 SPECTRIN 3.  
FT REPEAT 720 829 SPECTRIN 4.  
FT REPEAT 831 935 SPECTRIN 5.  
FT REPEAT 944 1047 SPECTRIN 6.  
FT REPEAT 1050 1156 SPECTRIN 7.  
FT REPEAT 1159 1265 SPECTRIN 8.  
FT REPEAT 1268 1369 SPECTRIN 9.  
FT REPEAT 1470 1570 SPECTRIN 10.  
FT REPEAT 1573 1678 SPECTRIN 11.  
FT REPEAT 1681 1782 SPECTRIN 12.  
FT REPEAT 1879 1981 SPECTRIN 13.  
FT REPEAT 2013 2103 SPECTRIN 14.  
FT REPEAT 2106 2210 SPECTRIN 15.  
FT REPEAT 2213 2320 SPECTRIN 16.  
FT REPEAT 2470 2572 SPECTRIN 17.  
FT REPEAT 2575 2681 SPECTRIN 18.  
FT REPEAT 2684 2797 SPECTRIN 19.  
FT REPEAT 2800 2902 SPECTRIN 20.  
FT REPEAT 2904 2926 SPECTRIN 21.  
FT REPEAT 2929 3035 SPECTRIN 22.  
FT DOMAIN 3050 3083 WW.

FT ZN PING 3302 3349 ZZ-TYPE.  
SQ SEQUENCE 3680 AA; 425650 MW; 539F1C9D72377872 CRC64;  
Query Match 42.2%; Score 43.5; DB 1; Length 3680;  
Best Local Similarity 55.6%; Pred. No. 3.5e+02;  
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;  
QY 3 LKYAQEELSNEVLPPPRK 20  
Db 697 VKHAQEELPP--PPPOK 711  
Search completed: May 11, 2004, 13:38:02  
Job time : 8.75 secs

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:28:26 ; Search time 32 Seconds  
(without alignments)  
197.199 Million cell updates/sec

Title: US-09-171-432A-41

Perfect score: 103

Sequence: 1 QRLKYAQEELSNEVLPPPRK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	100.0	82	12 Q8BDN1	Q8BDN1 hepatitis a
2	103	100.0	82	12 Q8BDN0	Q8BDN0 hepatitis a
3	103	100.0	82	12 Q8BDM9	Q8BDM9 hepatitis a
4	103	100.0	82	12 Q8BDM8	Q8BDM8 hepatitis a
5	103	100.0	82	12 Q8BDM7	Q8BDM7 hepatitis a
6	103	100.0	82	12 Q8BDM6	Q8BDM6 hepatitis a
7	103	100.0	82	12 Q8BDM5	Q8BDM5 hepatitis a
8	103	100.0	82	12 Q8BDM4	Q8BDM4 hepatitis a
9	103	100.0	82	12 Q8BDM3	Q8BDM3 hepatitis a
10	103	100.0	82	12 Q8BDM2	Q8BDM2 hepatitis a
11	103	100.0	82	12 Q8BDM1	Q8BDM1 hepatitis a
12	103	100.0	82	12 Q8BDM0	Q8BDM0 hepatitis a
13	103	100.0	82	12 Q8BDL9	Q8BDL9 hepatitis a
14	103	100.0	82	12 Q8BDL8	Q8BDL8 hepatitis a
15	103	100.0	82	12 Q8BDL7	Q8BDL7 hepatitis a
16	103	100.0	82	12 Q8BDL6	Q8BDL6 hepatitis a

90 103 100.0 116 12 Q9W7X7 hepatitis a  
91 103 100.0 116 12 Q9W7S7 hepatitis a  
92 103 100.0 116 12 Q71977 hepatitis a  
93 103 100.0 116 12 Q9W7T0 hepatitis a  
94 103 100.0 116 12 Q71975 hepatitis a  
95 103 100.0 116 12 Q71978 hepatitis a  
96 103 100.0 116 12 Q8B8K6 hepatitis a  
97 103 100.0 116 12 Q8B8K5 hepatitis a  
98 103 100.0 116 12 Q8B8K4 hepatitis a  
99 103 100.0 116 12 Q8B8K3 hepatitis a  
100 103 100.0 116 12 Q8B8K2 hepatitis a

# ALIGNMENTS

RESULT 1  
Q8BDN1 PRELIMINARY; PRT; 82 AA.  
AC Q8BDN1  
DT 01-MAR-2003 (TremBrel. 23, Created)  
DT 01-MAR-2003 (TremBrel. 23, Last sequence update)  
DT 01-MAR-2003 (TremBrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg283;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL; AF453451; AAN52259.1; -.  
FT NON\_TER 1  
FT NON\_TER 82  
SQ SEQUENCE 82 AA; 9505 MW; 8085191744E9C6E8 CRC64;

Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 61 QRLKYAQEELSNEVLPPPRK 80

RESULT 2  
Q8BDN0 PRELIMINARY; PRT; 82 AA.  
AC Q8BDN0  
DT 01-MAR-2003 (TremBrel. 23, Created)  
DT 01-MAR-2003 (TremBrel. 23, Last sequence update)  
DT 01-MAR-2003 (TremBrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg058;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL; AF453452; AAN52260.1; -.  
FT NON\_TER 1  
FT NON\_TER 82

SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 61 QRLKYAQEELSNEVLPPPRK 80

RESULT 3  
Q8BDM9 PRELIMINARY; PRT; 82 AA.  
AC Q8BDM9  
DT 01-MAR-2003 (TremBrel. 23, Created)  
DT 01-MAR-2003 (TremBrel. 23, Last sequence update)  
DT 01-MAR-2003 (TremBrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg987;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL; AF453453; AAN52261.1; -.  
FT NON\_TER 1  
FT NON\_TER 82  
SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;

Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 61 QRLKYAQEELSNEVLPPPRK 80

RESULT 4  
Q8BDM8 PRELIMINARY; PRT; 82 AA.  
AC Q8BDM8  
DT 01-MAR-2003 (TremBrel. 23, Created)  
DT 01-MAR-2003 (TremBrel. 23, Last sequence update)  
DT 01-MAR-2003 (TremBrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg441;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL; AF453454; AAN52262.1; -.  
FT NON\_TER 1  
FT NON\_TER 82  
SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;

Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 61 QRLKYAQEELSNEVLPPPRK 80

RESULT 5  
Q8BDM7 PRELIMINARY; PRT; 82 AA.

AC Q8BDM7; (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg799;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL; AF453455; AAN52263.1; --  
FT NON\_TER 1 1  
FT NON\_TER 82 82  
SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;

Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 61 QRLKYAQEELSNEVLPPPRK 80

RESULT 6  
Q8BDM6 PRELIMINARY; PRT; 82 AA.

AC Q8BDM6; (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg002;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL; AF453456; AAN52264.1; --  
FT NON\_TER 1 1  
FT NON\_TER 82 82  
SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;

Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 61 QRLKYAQEELSNEVLPPPRK 80

RESULT 7  
Q8BDM5 PRELIMINARY; PRT; 82 AA.

AC Q8BDM5; (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg139;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL; AF453457; AAN52265.1; --  
FT NON\_TER 1 1  
FT NON\_TER 82 82  
SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;

Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 61 QRLKYAQEELSNEVLPPPRK 80

RESULT 8  
Q8BDM4 PRELIMINARY; PRT; 82 AA.

AC Q8BDM4; (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg927;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL; AF453458; AAN52266.1; --  
FT NON\_TER 1 1  
FT NON\_TER 82 82  
SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;

Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 61 QRLKYAQEELSNEVLPPPRK 80

RESULT 9  
Q8BDM3 PRELIMINARY; PRT; 82 AA.

AC Q8BDM3;



01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Arg548;  
 RX MEDLINE=22198947; PubMed=12210404;  
 RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
 RT "Genetic characterization of hepatitis A virus isolates from Buenos  
 Aires, Argentina."  
 RL J. Med. Virol. 68:168-174(2002).  
 DR EMBL: AF453459; AAN52267.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 82  
 SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;  
 Query Match 100.0%; Score 103; DB 12; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 Db 61 QRLKYAQEELSNEVLPPPRK 80  
 RESULT 10  
 Q8BDM2 PRELIMINARY; PRT; 82 AA.  
 AC Q8BDM2;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Arg156;  
 RX MEDLINE=22198947; PubMed=12210404;  
 RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
 RT "Genetic characterization of hepatitis A virus isolates from Buenos  
 Aires, Argentina."  
 RL J. Med. Virol. 68:168-174(2002).  
 DR EMBL: AF453460; AAN52268.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 82  
 SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;  
 Query Match 100.0%; Score 103; DB 12; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 Db 61 QRLKYAQEELSNEVLPPPRK 80  
 RESULT 11  
 Q8BDM1 PRELIMINARY; PRT; 82 AA.  
 AC Q8BDM1;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.

Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Arg642;  
 RX MEDLINE=22198947; PubMed=12210404;  
 RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
 RT "Genetic characterization of hepatitis A virus isolates from Buenos  
 Aires, Argentina."  
 RL J. Med. Virol. 68:168-174(2002).  
 DR EMBL: AF453461; AAN52269.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 82  
 SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;  
 Query Match 100.0%; Score 103; DB 12; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 Db 61 QRLKYAQEELSNEVLPPPRK 80  
 RESULT 12  
 Q8BDM0 PRELIMINARY; PRT; 82 AA.  
 AC Q8BDM0;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Arg784;  
 RX MEDLINE=22198947; PubMed=12210404;  
 RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
 RT "Genetic characterization of hepatitis A virus isolates from Buenos  
 Aires, Argentina."  
 RL J. Med. Virol. 68:168-174(2002).  
 DR EMBL: AF453462; AAN52270.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 82  
 SQ SEQUENCE 82 AA; 9503 MW; 8083C4D698E9C6E8 CRC64;  
 Query Match 100.0%; Score 103; DB 12; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRLKYAQEELSNEVLPPPRK 20  
 Db 61 QRLKYAQEELSNEVLPPPRK 80  
 RESULT 13  
 Q8BDL9 PRELIMINARY; PRT; 82 AA.  
 AC Q8BDL9;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=Arg9818;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL: AF453463; AAN52271.1; -.  
FT NON TER 1  
FT NON TER 82  
SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 14  
Q8BDL8 PRELIMINARY; PRT; 82 AA.  
AC Q8BDL8;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg973;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL: AF453464; AAN52272.1; -.  
FT NON TER 1  
FT NON TER 82  
SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 15  
Q8BDL7 PRELIMINARY; PRT; 82 AA.  
AC Q8BDL7;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg143;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."

RL J. Med. Virol. 68:168-174(2002).  
DR EMBL: AF453465; AAN52273.1; -.  
FT NON TER 1  
FT NON TER 82  
SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 16  
Q8BDL6 PRELIMINARY; PRT; 82 AA.  
AC Q8BDL6;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg081;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL: AF453466; AAN52274.1; -.  
FT NON TER 1  
FT NON TER 82  
SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 82;  
Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 17  
Q8BDL5 PRELIMINARY; PRT; 82 AA.  
AC Q8BDL5;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Arg578;  
RX MEDLINE=22198947; PubMed=12210404;  
RA Mbayed V.A., Sookoian S., Alfonso V., Campos R.H.;  
RT "Genetic characterization of hepatitis A virus isolates from Buenos Aires, Argentina."  
RL J. Med. Virol. 68:168-174(2002).  
DR EMBL: AF453467; AAN52275.1; -.  
FT NON TER 1  
FT NON TER 82  
SQ SEQUENCE 82 AA; 9521 MW; 808504D698E9C6E8 CRC64;



FT NON TER 1 1  
FT NON TER 94 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 22  
Q9ENU0 PRELIMINARY; PRT; 94 AA.  
AC Q9ENU0;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 30;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038297; BAB11835.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
FT NON TER 1 1  
FT NON TER 94 94  
SQ SEQUENCE 94 AA; 10844 MW; 9AF8A263BB895FAC CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 23  
Q9ENS8 PRELIMINARY; PRT; 94 AA.  
AC Q9ENS8;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Kantou 69;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038303; BAB11847.1; -.  
DR PIR; PQ0427; PQ0427.

DR PIR; PQ0428; PQ0428.  
FT NON TER 1 1  
FT NON TER 94 94  
SQ SEQUENCE 94 AA; 10842 MW; 874E7991BB895FB9 CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 24  
Q9ENT8 PRELIMINARY; PRT; 94 AA.  
AC Q9ENT8;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ACC 1;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038299; BAB11837.1; -.  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
FT NON TER 1 1  
FT NON TER 94 94  
SQ SEQUENCE 94 AA; 10862 MW; 9AF8A263BB8C4BAC CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
Db 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 25  
Q9ENU2 PRELIMINARY; PRT; 94 AA.  
AC Q9ENU2;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 28;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
infection: Prolonged HAV viremia and mild liver injury.";  
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038295; BAB11833.1; -.

DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 26  
Q9ENT3 PRELIMINARY; PRT; 94 AA.  
ID Q9ENT3  
AC Q9ENT3;  
DT 01-MAR-2001 (TREMELrel. 16, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Polypeptide (fragment)  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Kantou 45;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.,  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038304; BAB11842.1; --  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10828 MW; 74184C635B86E1A6 CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 27  
Q9ENT1 PRELIMINARY; PRT; 94 AA.  
ID Q9ENT1  
AC Q9ENT1;  
DT 01-MAR-2001 (TREMELrel. 16, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Polypeptide (fragment)  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Kantou 49;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.,  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)

RT infection: Prolonged HAV viremia and mild liver injury."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038306; BAB11844.1; --  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10842 MW; 741850915B86E1A6 CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 28  
Q9ENT0 PRELIMINARY; PRT; 94 AA.  
ID Q9ENT0  
AC Q9ENT0;  
DT 01-MAR-2001 (TREMELrel. 16, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Polypeptide (fragment)  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Kantou 53;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.,  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038307; BAB11845.1; --  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
FT NON\_TER 1  
SQ SEQUENCE 94 AA; 10858 MW; 9AF8BE91BB895FAC CRC64;  
Query Match 100.0%; Score 103; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 61 QRLKYAQEELSNEVLPPPRK 80  
RESULT 29  
Q9ENU1 PRELIMINARY; PRT; 94 AA.  
ID Q9ENU1  
AC Q9ENU1;  
DT 01-MAR-2001 (TREMELrel. 16, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Polypeptide (fragment)  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 29;

```

RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,
RA Yasuoka A., Oka S.;
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)
RT infection: Prolonged HAV viremia and mild liver injury.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB038296; BAB11834.1; -.
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
FT NON_TER 1
FT NON_TER 94
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE31BB8C4BAC CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ORLYAQEELSNEVLPPRK 20
Db 61 ORLYAQEELSNEVLPPRK 80

RESULT 30
Q9ENT4 PRELIMINARY; PRT; 94 AA.
AC Q9ENT4;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
CX NCBI_TaxID=12092;
RN [1]_TaxID=12092;
RP SEQUENCE FROM N.A.
RC STRAIN=Kantou 44;
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,
RA Yasuoka A., Oka S.;
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)
RT infection: Prolonged HAV viremia and mild liver injury.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB038303; BAB11841.1; -.
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
FT NON_TER 1
FT NON_TER 94
SQ SEQUENCE 94 AA; 10828 MW; 74184C63BB86E1A6 CRC64;

Query Match 100.0%; Score 103; DB 12; Length 94;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ORLYAQEELSNEVLPPRK 20
Db 61 ORLYAQEELSNEVLPPRK 80

Search completed: May 11, 2004, 13:40:24
Job time : 33 secs

```

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:34:17 ; Search time 13.5 Seconds  
(without alignments)  
76.483 Million cell updates/sec

Title: US-09-171-432A-41

Perfect score: 103

Sequence: 1 QRLKYAQEELSNEVLPPPRK 20

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

- 1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCFUS COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	1091	6	Patent No. 5516630-2
2	103	100.0	2227	3	US-08-475-886-2
3	103	100.0	2227	3	US-08-475-886-5
4	103	100.0	2227	3	US-08-397-232-2
5	103	100.0	2227	3	US-08-397-232-4
6	103	100.0	2227	3	US-09-171-387-2
7	103	100.0	2227	4	US-09-653-499-2
8	103	100.0	2227	4	US-09-653-499-6
9	103	100.0	2227	4	US-10-104-966-12
10	103	100.0	2227	4	US-10-135-988-2
11	103	100.0	2227	4	US-10-135-988-6
12	98	95.1	2227	3	US-08-475-886-4
13	98	95.1	2227	4	US-09-653-499-4
14	98	95.1	2227	4	US-10-135-988-4
15	94	91.3	839	1	US-09-252-991A-18339
16	44	42.7	914	1	US-08-484-105-2
17	44	42.7	914	1	US-08-484-106-2
18	44	42.7	914	1	US-08-241-853-15
19	43.5	42.2	113	1	US-08-850-917-15
20	43.5	42.2	113	2	US-08-850-917-15
21	43	41.7	209	4	US-09-543-681A-7657
22	43	41.7	390	4	US-09-308-003-12
23	43	41.7	414	4	US-09-252-991A-22296
24	43	41.7	585	2	US-08-453-848-11
25	43	41.7	585	3	US-09-169-027-11
26	43	41.7	586	2	US-08-453-848-19
27	43	41.7	586	3	US-09-169-027-19

ALIGNMENTS

RESULT 1  
Patent No. 5516630  
; APPLICANT: TICEHURST, JOHN R.; BALTIMORE, DAVID; FEINSTEIN, STEPHEN M.; PURCELL, ROBERT H.; RACANIELLO, VINCENT R.; BARCUDY, BAHRIG M.  
; TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/788,262  
; FILING DATE: 06-NOV-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 256,135  
; FILING DATE: 06-OCT-1988  
; APPLICATION NUMBER: 654,942  
; FILING DATE: 27-SEP-1984  
; APPLICATION NUMBER: 537,911  
; FILING DATE: 30-SEP-1983  
; SEQ ID NO: 2:  
; LENGTH: 1091  
5516630-2

Query Match 100.0%; Score 103; DB 6; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 6.4e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ORLKYAQEELSNEVLPPPRK 20  
Db 1047 ORLKYAQEELSNEVLPPPRK 1066

RESULT 2  
US-08-475-886-2  
; Sequence 2, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-08-475-886-2

Query Match 100.0%; Score 103; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ORLKYAQEELSNEVLPPPRK 20  
Db 810 ORLKYAQEELSNEVLPPPRK 829  
RESULT 3  
US-08-475-886-6  
; Sequence 6, Application US/08475886A

Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-08-475-886-6

Query Match 100.0%; Score 103; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ORLKYAQEELSNEVLPPPRK 20  
Db 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 4  
US-08-397-232-2  
; Sequence 2, Application US/08397232A  
; Patent No. 6180110  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US1  
; CURRENT APPLICATION NUMBER: US/08/397,232A  
; CURRENT FILING DATE: 1995-04-17  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: PCT/US93/08610  
; EARLIER FILING DATE: 1993-09-17  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-08-397-232-2

Query Match 100.0%; Score 103; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ORLKYAQEELSNEVLPPPRK 20  
Db 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 5  
US-08-397-232-4  
; Sequence 4, Application US/08397232A  
; Patent No. 6180110  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H



; APPLICANT: D'HONDT, ERIC  
 ; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
 ; FILE REFERENCE: 20264262US1  
 ; CURRENT APPLICATION NUMBER: US/08/397,232A  
 ; CURRENT FILING DATE: 1995-04-17  
 ; EARLIER APPLICATION NUMBER: 07/947,338  
 ; EARLIER FILING DATE: 1992-09-18  
 ; EARLIER APPLICATION NUMBER: PCT/US93/08610  
 ; EARLIER FILING DATE: 1993-09-17  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 4  
 ; LENGTH: 2227  
 ; TYPE: PRT  
 ; ORGANISM: Attenuated (4380) HAV, strain HM-175  
 US-08-397-232-4

Query Match 100.0%; Score 103; DB 3; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPRK 20  
 Db 810 QRLKYAQEELSNEVLPPRK 829

RESULT 6  
 US-09-171-387-2  
 ; Sequence 2, Application US/09171387  
 ; Patent No. 6280734  
 ; GENERAL INFORMATION:  
 ; APPLICANT: RAYCHAUDHURI, GOPA;  
 ; EMERSON, SUZANNE, U.;  
 ; PURCELL, ROBERT, H.  
 ; TITLE OF INVENTION: SIMIAN-HUMAN HAV  
 ; HAVING A CHIMERIC 2C PROTEIN  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
 ; STREET: 345 PARK AVENUE  
 ; CITY: NEW YORK  
 ; STATE: NEW YORK  
 ; COUNTRY: USA  
 ; ZIP: 10154

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: FLOPPY DISK  
 ; COMPUTER: IBM PC COMPATIBLE  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: MICROSOFT WORD 97  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/171,387  
 ; FILING DATE: 24-Mar-1999  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US97/06506  
 ; FILING DATE: 18-APR-1997  
 ; APPLICATION NUMBER: US60/015,642  
 ; FILING DATE: 19-APR-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: William S. Feller  
 ; REGISTRATION NUMBER: 26,728  
 ; REFERENCE/DOCKET NUMBER: 2026-4229US1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 758-4800  
 ; TELEFAX: (212) 751-6849  
 ; INFORMATION FOR SEQ ID NO: 2  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2227 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 US-09-171-387-2

Query Match 100.0%; Score 103; DB 3; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPRK 20  
 Db 810 QRLKYAQEELSNEVLPPRK 829

RESULT 7  
 US-09-653-499-2  
 ; Sequence 2, Application US/09653499  
 ; Patent No. 6423318  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FUNKHOUSER, ANN W  
 ; EMERSON, SUZANNE U  
 ; PURCELL, ROBERT H  
 ; D'HONDT, ERIC  
 ; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
 ; FILE REFERENCE: 20264262US2  
 ; CURRENT APPLICATION NUMBER: US/09/653,499  
 ; CURRENT FILING DATE: 2000-08-31  
 ; PRIOR APPLICATION NUMBER: 08/475,886  
 ; PRIOR FILING DATE: 1999-08-09  
 ; PRIOR APPLICATION NUMBER: 07/947,338  
 ; PRIOR FILING DATE: 1992-09-18  
 ; PRIOR APPLICATION NUMBER: 08/397,232  
 ; PRIOR FILING DATE: 1995-03-10  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 2227  
 ; TYPE: PRT  
 ; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
 US-09-653-499-2

Query Match 100.0%; Score 103; DB 4; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLPPRK 20  
 Db 810 QRLKYAQEELSNEVLPPRK 829

RESULT 8  
 US-09-653-499-6  
 ; Sequence 6, Application US/09653499  
 ; Patent No. 6423318  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FUNKHOUSER, ANN W  
 ; EMERSON, SUZANNE U  
 ; PURCELL, ROBERT H  
 ; D'HONDT, ERIC  
 ; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
 ; FILE REFERENCE: 20264262US2  
 ; CURRENT APPLICATION NUMBER: US/09/653,499  
 ; CURRENT FILING DATE: 2000-08-31  
 ; PRIOR APPLICATION NUMBER: 08/475,886  
 ; PRIOR FILING DATE: 1999-08-09  
 ; PRIOR APPLICATION NUMBER: 07/947,338  
 ; PRIOR FILING DATE: 1992-09-18  
 ; PRIOR APPLICATION NUMBER: 08/397,232  
 ; PRIOR FILING DATE: 1995-03-10  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 2227  
 ; TYPE: PRT  
 ; ORGANISM: Attenuated (4380) HAV, strain HM-175  
 US-09-653-499-6

Query Match 100.0%; Score 103; DB 4; Length 2227;

Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
|||||  
Db 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 9  
US-10-104-966-12  
; Sequence 12, Application US/10104966  
; Patent No. 6680059  
; GENERAL INFORMATION:  
; APPLICANT: Matti Sallberg  
; APPLICANT: Catharina Hultgren  
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND  
; FILE REFERENCE: TRIPEP.23AUSC1  
; CURRENT APPLICATION NUMBER: US/10/104,966  
; CURRENT FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: 09/705,547  
; PRIOR FILING DATE: 2000-11-03  
; PRIOR APPLICATION NUMBER: 60/229,175  
; PRIOR FILING DATE: 2000-08-29  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Hepatitis A virus sequence  
US-10-104-966-12

Query Match 100.0%; Score 103; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
|||||  
Db 810 QRLKYAQEELSNEVLPPPRK 829

RESULT 10  
US-10-135-988-2  
; Sequence 2, Application US/10135988  
; Patent No. 6680060  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-10-135-988-2

Query Match 100.0%; Score 103; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
|||||

Db 810 QRLKYAQEELSNEVLPPPRK 829  
|||||

RESULT 11  
US-10-135-988-6  
; Sequence 6, Application US/10135988  
; Patent No. 6680060  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-10-135-988-6

Query Match 100.0%; Score 103; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
|||||

Db 810 QRLKYAQEELSNEVLPPPRK 829  
|||||

RESULT 12  
US-08-475-886-4  
; Sequence 4, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175  
US-08-475-886-4

Query Match 95.1%; Score 98; DB 3; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 8.5e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVLPPPRK 20  
|||||

Db 810 QRLKYAQEELSNEVLPPPRK 829  
|||||

RESULT 13

US-09-653-499-4  
; Sequence 4, Application US/09653499  
; Patent No. 6423318  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/09/653,499  
; CURRENT FILING DATE: 2000-08-31  
; PRIOR APPLICATION NUMBER: 08/475,886  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175  
US-09-653-499-4  
Query Match 95.1%; Score 98; DB 4; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 8.5e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 810 QRLKYAQEELSNEVLPPPRK 829  
RESULT 14  
US-10-135-988-4  
; Sequence 4, Application US/10135988  
; Patent No. 6680060  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175  
US-10-135-988-4  
Query Match 95.1%; Score 98; DB 4; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 8.5e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 810 QRLKYAQEELSNEVLPPPRK 829  
RESULT 15  
US-08-087-016-2  
; Sequence 2, Application US/08087016  
; Patent No. 5430135

GENERAL INFORMATION:  
; APPLICANT: NAINAN, OMANA V.  
; APPLICANT: MARGOLIS, HAROLD S.  
; APPLICANT: ROBERTSON, BETTY H.  
; APPLICANT: BRINTON, MARGO H.  
; APPLICANT: EBERT, JAMES W.  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1615 L Street N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/087,016  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/678,828  
; FILING DATE: 03-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26,581  
; REFERENCE/DOCKET NUMBER: WTS/5683/83834/SRL  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 839 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-087-016-2  
Query Match 91.3%; Score 94; DB 1; Length 839;  
Best Local Similarity 85.0%; Pred. No. 1.3e-06;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPPRK 20  
DB 809 QRLKYAQEELSNEVLPPPRK 828  
RESULT 16  
US-09-252-991A-18339  
; Sequence 18339, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 18339  
; LENGTH: 180  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18339

Query Match 42.7%; Score 44; DB 4; Length 180;  
Best Local Similarity 47.4%; Pred. No. 17;  
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPRK 19  
Db 29 RRLRYAAQLRQRLVLPQR 47  
RESULT 17  
US-08-484-105-2  
Sequence 2, Application US/08484105  
Patent No. 5589341  
GENERAL INFORMATION:  
APPLICANT: STILLMAN, Bruce  
APPLICANT: BELL, Stephen P  
APPLICANT: KOBAYASHI, Ryuji  
APPLICANT: RINE, Jasper  
APPLICANT: FOSS, Margit  
APPLICANT: McNALLY, Francis J  
APPLICANT: LAURENSEN, Patricia  
APPLICANT: HERSKOWITZ, Ira  
APPLICANT: LI, Joachim J  
APPLICANT: GAVIN, Kimberly  
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,105  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard Aron  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 914 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-105-2  
Query Match 42.7%; Score 44; DB 1; Length 914;  
Best Local Similarity 30.0%; Pred. No. 1e+02;  
Matches 6; Conservative 8; Mismatches 6; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPRK 20  
Db 310 RKIKGKDDIDASVQPPPK 329  
RESULT 19  
US-08-241-853-15  
Sequence 15, Application US/08241853  
Patent No. 5693488  
GENERAL INFORMATION:  
APPLICANT: Fang, Kathy S.  
APPLICANT: Hanfusa, Hidesaburo  
TITLE OF INVENTION: A TRANSMEMBRANE TYROSINE PHOSPHATASE  
TITLE OF INVENTION: AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:

GENERAL INFORMATION:  
APPLICANT: STILLMAN, Bruce  
APPLICANT: BELL, Stephen P  
APPLICANT: KOBAYASHI, Ryuji  
APPLICANT: RINE, Jasper  
APPLICANT: FOSS, Margit  
APPLICANT: McNALLY, Francis J  
APPLICANT: LAURENSEN, Patricia  
APPLICANT: HERSKOWITZ, Ira  
APPLICANT: LI, Joachim J  
APPLICANT: GAVIN, Kimberly  
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,106  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard Aron  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 914 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-106-2  
Query Match 42.7%; Score 44; DB 1; Length 914;  
Best Local Similarity 30.0%; Pred. No. 1e+02;  
Matches 6; Conservative 8; Mismatches 6; Indels 0; Gaps 0;  
QY 1 QRLKYAQEELSNEVLPPRK 20  
Db 310 RKIKGKDDIDASVQPPPK 329  
RESULT 19  
US-08-241-853-15  
Sequence 15, Application US/08241853  
Patent No. 5693488  
GENERAL INFORMATION:  
APPLICANT: Fang, Kathy S.  
APPLICANT: Hanfusa, Hidesaburo  
TITLE OF INVENTION: A TRANSMEMBRANE TYROSINE PHOSPHATASE  
TITLE OF INVENTION: AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/241,853  
FILING DATE: 12-MAY-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-078  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 113 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-241-853-15  
Query Match 42.2%; Score 43.5; DB 1; Length 113;  
Best Local Similarity 55.6%; Pred. No. 13;  
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;  
QY 3 LKYAQEELSNEVLPPPRK 20  
DB 7 VKHAQEELPP---PPQK 21  
RESULT 20  
US-08-850-917-15  
Sequence 15, Application US/08850917  
Patent No. 5854045  
GENERAL INFORMATION:  
APPLICANT: Fang, Kathy S.  
APPLICANT: Hanafusa, Hidesaburo  
TITLE OF INVENTION: A TRANSMEMBRANE TYROSINE PHOSPHATASE  
TITLE OF INVENTION: AND METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/850,917  
FILING DATE: 02-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/241,853  
FILING DATE: 12-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-078  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521

INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 113 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-850-917-15  
Query Match 42.2%; Score 43.5; DB 2; Length 113;  
Best Local Similarity 55.6%; Pred. No. 13;  
Matches 10; Conservative 3; Mismatches 2; Indels 3; Gaps 1;  
QY 3 LKYAQEELSNEVLPPPRK 20  
DB 7 VKHAQEELPP---PPQK 21  
RESULT 21  
US-09-543-681A-7657  
Sequence 7657, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 7657  
LENGTH: 209  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-7657  
Query Match 41.7%; Score 43; DB 4; Length 209;  
Best Local Similarity 61.5%; Pred. No. 29;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
QY 4 KYAQEELSNEVL 16  
DB 100 KFALEKFSNELLP 112  
RESULT 22  
US-09-308-003-12  
Sequence 12, Application US/09308003  
Patent No. 6326170  
GENERAL INFORMATION:  
APPLICANT: Burnham, Martin K. R.  
APPLICANT: Lonetto, Michael A.  
APPLICANT: Warren, Patrick V.  
TITLE OF INVENTION: NOVEL PROKARYOTIC POLYNUCLEOTIDES,  
TITLE OF INVENTION: POLYPEPTIDES AND THEIR USES  
FILE REFERENCE: GM10093  
CURRENT APPLICATION NUMBER: US/09/308,003  
CURRENT FILING DATE: 1999-05-10  
EARLIER APPLICATION NUMBER: 60/058,710  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 52  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 390  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-308-003-12  
Query Match 41.7%; Score 43; DB 4; Length 390;

Best Local Similarity 57.1%; Pred. No. 58;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 LKVAQEELSNEVL 16  
Db 368 IKVALKEASNEIP 361

RESULT 23  
US-09-252-991A-22296  
; Sequence 22296, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22296  
; LENGTH: 414  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22296

Query Match 41.7%; Score 43; DB 4; Length 414;  
Best Local Similarity 55.6%; Pred. No. 62;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 QRLKVAQEELSNEVLPP 18  
Db 342 QRLDYHAGELDLLELPP 359

RESULT 24  
US-08-453-848-11  
; Sequence 11, Application US/08453848  
; Patent No. 5858368  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Gale Eugene  
; APPLICANT: Volvovitz, Franklin  
; APPLICANT: Wilkinson, Bethanie Eident  
; APPLICANT: Voznesensky, Andrei I.  
; APPLICANT: Hackett, Craig Stanway  
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: GA  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,848  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/120,607  
; FILING DATE: 13-SEPT-1993  
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: MGS101CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404)-873-8794  
; TELEFAX: (404)-873-8795  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 585 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Influenza virus  
; INDIVIDUAL ISOLATE: B/Panama/45/90 rHA  
; FEATURE:  
; NAME/KEY: HA signal peptide  
; LOCATION: 1 to 17  
; FEATURE:  
; NAME/KEY: mature rHA  
; LOCATION: 18 to 568  
US-08-453-848-11

Query Match 41.7%; Score 43; DB 2; Length 585;  
Best Local Similarity 60.0%; Pred. No. 90;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QRLKVAQEELSNEVL 15  
Db 426 QRLSGAMDELHNEIL 440

RESULT 25  
US-09-169-027-11  
; Sequence 11, Application US/09169027  
; Patent No. 6245532  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Gale Eugene  
; APPLICANT: Volvovitz, Franklin  
; APPLICANT: Wilkinson, Bethanie Eident  
; APPLICANT: Voznesensky, Andrei I.  
; APPLICANT: Hackett, Craig Stanway  
; TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: GA  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/169,027  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,848  
; FILING DATE: 30-MAY-1995  
; APPLICATION NUMBER: 08/120,607  
; FILING DATE: 13-SEPT-1993  
; ATTORNEY/AGENT INFORMATION:

```

NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: MGS101CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)-873-8794
TELEFAX: (404)-873-8795
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 585 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Influenza virus
INDIVIDUAL ISOLATE: B/Panama/45/90 rHA
FEATURE:
NAME/KEY: HA signal peptide
LOCATION: 1 to 17
FEATURE:
NAME/KEY: mature rHA
LOCATION: 18 to 568
US-09-169-027-11

Query Match 41.7%; Score 43; DB 3; Length 585;
Best Local Similarity 60.0%; Pred. No. 90;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVL 15
DB 426 QRLSGAMDELHNEIL 440

```

426 QRLSGAMDELHNEIL 440  
 DB  
 RESULT 26  
 US-08-453-848-19  
 ? Sequence 19, Application US/08453848  
 ? Patent No. 5858368  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Smith, Gale Eugene  
 ? APPLICANT: Volvevitz, Franklin  
 ? APPLICANT: Wilkinson, Bethanie Bident  
 ? APPLICANT: Voznesensky, Andrei I.  
 ? APPLICANT: Hackett, Craig Stanway  
 ? TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
 ? TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES  
 ? NUMBER OF SEQUENCES: 31  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Patrea L. Pabst  
 ? STREET: 2800 One Atlantic Center  
 ? STREET: 1201 West Peachtree Street  
 ? CITY: Atlanta  
 ? STATE: GA  
 ? COUNTRY: USA  
 ? ZIP: 30309-3450  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: Floppy disk  
 ? COMPUTER: IBM PC compatible  
 ? OPERATING SYSTEM: PC-DOS/MS-DOS  
 ? SOFTWARE: Patent In Release #1.0, Version #1.25  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: US/08/453,848  
 ? FILING DATE: 30-MAY-1995  
 ? CLASSIFICATION: 435  
 ? PRIOR APPLICATION DATA:  
 ? APPLICATION NUMBER: 08/120,607  
 ? FILING DATE: 13-SEPT-1993  
 ? CLASSIFICATION: 435  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Pabst, Patrea L.  
 ? REGISTRATION NUMBER: 31,284





INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 589 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Influenza virus  
INDIVIDUAL ISOLATE: B/Netherlands/13/94 rHA  
FEATURE:  
NAME/KEY: AcNPV 61K protein signal sequence  
LOCATION: 1 to 18  
FEATURE:  
NAME/KEY: mature rHA  
LOCATION: 19 to 571  
US-09-169-027-13

Query Match 41.7%; Score 43; DB 3; Length 589;  
Best Local Similarity 60.0%; Pred. No. 90;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVL 15  
||| | :||| :|||  
Db 430 QRLGAMDELHNEIL 444

RESULT 30  
US-08-453-848-17  
Sequence 17, Application US/08453848  
Patent No. 5858368  
GENERAL INFORMATION:  
APPLICANT: Smith, Gale Eugene  
APPLICANT: Volvovitz, Franklin  
APPLICANT: Wilkinson, Bethanie Eident  
APPLICANT: Voznesensky, Andrei I.  
APPLICANT: Hackett, Craig Stanway  
TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: GA  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,848  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/120,607  
FILING DATE: 13-SEPT-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: MGS101CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404)-873-8794  
TELEFAX: (404)-873-8795  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:

LENGTH: 592 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Influenza virus  
INDIVIDUAL ISOLATE: B/Shanghai/4/94 rHA  
FEATURE:  
NAME/KEY: AcNPV 61K protein signal peptide  
LOCATION: 1 to 18  
FEATURE:  
NAME/KEY: mature rHA  
LOCATION: 19 to 574  
US-08-453-848-17

Query Match 41.7%; Score 43; DB 2; Length 592;  
Best Local Similarity 60.0%; Pred. No. 91;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QRLKYAQEELSNEVL 15  
||| | :||| :|||  
Db 433 QRLGAMDELHNEIL 447

Search completed: May 11, 2004, 13:42:21  
Job time : 14.5 secs

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:37:37 ; Search time 34.25 Seconds  
(without alignments)  
162.083 Million cell updates/sec

Title: US-09-171-432A-41

Perfect score: 103

Sequence: 1 QRLKYAQEELSNEVLPPPRK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications RA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	20	10 US-09-171-432A-41	Sequence 41, Appl
2	103	100.0	25	10 US-09-171-432A-47	Sequence 47, Appl
3	103	100.0	352	14 US-10-272-459-45	Sequence 45, Appl
4	103	100.0	836	14 US-10-272-459-40	Sequence 40, Appl
5	103	100.0	980	14 US-10-272-459-41	Sequence 41, Appl
6	103	100.0	2227	9 US-09-929-955-12	Sequence 12, Appl
7	103	100.0	2227	13 US-10-104-966-12	Sequence 12, Appl
8	103	100.0	2227	13 US-10-135-988-2	Sequence 2, Appl
9	103	100.0	2227	13 US-10-135-988-6	Sequence 6, Appl
10	98	95.1	2227	13 US-10-135-988-4	Sequence 4, Appl
11	52	50.5	1193	15 US-10-369-493-5713	Sequence 5713, Ap
12	51	49.5	1240	15 US-10-369-493-4031	Sequence 4031, Ap
13	48	46.6	63	12 US-10-424-599-254867	Sequence 254867,
14	48	46.6	442	12 US-10-335-977-5788	Sequence 5788, Ap
15	48	46.6	444	12 US-10-335-977-5789	Sequence 5789, Ap

```

RESULT 2
US-09-171-432A-47
; Sequence 47, Application US/09171432A
; Publication No. US20030187184A1
; GENERAL INFORMATION:
; APPLICANT: Fields, Howard A.
; APPLICANT: Khudyakov, Yury E.
; TITLE OF INVENTION: Antigenically Reactive Regions of the
; TITLE OF INVENTION: Hepatitis A Virus Polyprotein
; NUMBER OF SEQUENCES: 88
;

```

## STREET: 3424 PE

COUNTRY: USA  
ZIP: 30326  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/171.432A  
FILING DATE: 23-NOV-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Greene, Jamie L.

REGISTRATION NUMBER: 32,467  
REFERENCE/DOCKET NUMBER: 03063-0231US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 949-2400  
TELEFAX: (404) 949-2499  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

```

;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..25
; OTHER INFORMATION: /label= YK-1665
;
US-09-171-432A-47

```

```

Query Match      100.0%; Score 103; DB 10; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps

```

US-10-272-459-45  
; Sequence 45, Application US/10272459  
; Publication NO. US20030124517A1  
; GENERAL INFORMATION:

```

1  APPLICANT: PICHUANYES, SERGIO
2  TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT
3  TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT
4  TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT
5  TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT
6  TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT
7  FILE REFERENCE: PPI/955.002 / 2301-17955
8  CURRENT APPLICATION NUMBER: US/10/272,459
9  CURRENT FILING DATE: 2002-10-15
10 NUMBER OF SEQ ID NOS: 50
11 SOFTWARE: Patentin Ver. 2.0

```

```

; SEQ ID NO 45
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein
; OTHER INFORMATION: of 38.8 kDa
US-10-272-459-45

Query Match      100.0%; Score 103; DB 14; Length 352;
Best Local Similarity 100.0%; Pred. No. 4.9e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20
   |||||
DB 182 ORLKYAQEELSNEVLPPPRK 201

RESULT 4
US-10-272-459-40
; Sequence 40, Application US/10272459
; Publication No. US20030124517A1
; GENERAL INFORMATION:
; APPLICANT: PICHUANTES, Sergio
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: PPI7955.002 / 2301-17955
; CURRENT APPLICATION NUMBER: US/10/272,459
; CURRENT FILING DATE: 2002-10-15
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 40
; LENGTH: 836
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein
; OTHER INFORMATION: of 94 kDa
US-10-272-459-40

Query Match      100.0%; Score 103; DB 14; Length 836;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20
   |||||
DB 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 5
US-10-272-459-41
; Sequence 41, Application US/10272459
; Publication No. US20030124517A1
; GENERAL INFORMATION:
; APPLICANT: PICHUANTES, Sergio
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: PPI7955.002 / 2301-17955
; CURRENT APPLICATION NUMBER: US/10/272,459
; CURRENT FILING DATE: 2002-10-15
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 41
; LENGTH: 980
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein
; OTHER INFORMATION: of 115.5 kDa
US-10-272-459-41

Query Match      100.0%; Score 103; DB 14; Length 980;
Best Local Similarity 100.0%; Pred. No. 1.5e-06;

```

```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20
   |||||
DB 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 6
US-09-929-955-12
; Sequence 12, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUS2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis A virus sequence
US-09-929-955-12

Query Match      100.0%; Score 103; DB 9; Length 2227;
Best Local Similarity 100.0%; Pred. No. 3.6e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORLKYAQEELSNEVLPPPRK 20
   |||||
DB 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 7
US-10-104-966-12
; Sequence 12, Application US/10104966
; Publication No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUSCI
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis A virus sequence
US-10-104-966-12

Query Match      100.0%; Score 103; DB 13; Length 2227;
Best Local Similarity 100.0%; Pred. No. 3.6e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 ORLKYAQEELSNEVLPPPRK 20  
|||||  
Db 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 8  
US-10-135-988-2  
; Sequence 2, Application US/10135988  
; Publication No. US20020176869A1  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175

US-10-135-988-2

Query Match 100.0%; Score 103; DB 13; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3.6e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ORLKYAQEELSNEVLPPPRK 20  
|||||  
Db 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 9  
US-10-135-988-6  
; Sequence 6, Application US/10135988  
; Publication No. US20020176869A1  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated (4380) HAV, strain HM-175

Query Match 100.0%; Score 103; DB 13; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3.6e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ORLKYAQEELSNEVLPPPRK 20  
|||||  
Db 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 10  
US-10-135-988-4  
; Sequence 4, Application US/10135988  
; Publication No. US20020176869A1  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US3  
; CURRENT APPLICATION NUMBER: US/10/135,988  
; CURRENT FILING DATE: 2002-04-29  
; PRIOR APPLICATION NUMBER: 07/947,338  
; PRIOR FILING DATE: 1992-09-18  
; PRIOR APPLICATION NUMBER: 08/397,232  
; PRIOR FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175

Query Match 95.1%; Score 98; DB 13; Length 2227;  
Best Local Similarity 95.0%; Pred. No. 2e-05; 0; Indels 0; Gaps 0;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ORLKYAQEELSNEVLPPPRK 20  
|||||  
Db 810 ORLKYAQEELSNEVLPPPRK 829

RESULT 11  
US-10-369-493-5713  
; Sequence 5713, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 5713  
; LENGTH: 1193  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans

Query Match 50.5%; Score 52; DB 15; Length 1193;  
Best Local Similarity 62.5%; Pred. No. 65;  
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ORLKYAQEELSNEVL 16  
|||||  
Db 340 QRIKYAREILOKELP 355

RESULT 12  
US-10-369-493-4031  
; Sequence 4031, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:



APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragoras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 5789:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 444 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...444  
SEQUENCE DESCRIPTION: SEQ ID NO: 5789:  
US-10-335-977-5789  
Query Match 46.6%; Score 48; DB 12; Length 444;  
Best Local Similarity 75.0%; Pred. No. 87;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 9 ELSNEVLPPPK 20  
DB 416 EADNEELPPPK 427  
RESULT 16  
US-10-424-599-234250  
Sequence 234250, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 234250  
LENGTH: 259  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_53552C.1.pep  
US-10-424-599-234250  
Query Match 45.6%; Score 47; DB 12; Length 259;  
Best Local Similarity 45.5%; Pred. No. 88;  
Matches 10; Conservative 4; Mismatches 6; Indels 2; Gaps 1;  
QY 1 QRLKYAQEELSN--EVLPPPK 20  
DB 106 EMLPYLNEKLKRNVEVPPPK 127  
RESULT 17  
US-10-425-114-44229  
Sequence 44229, Application US/10425114  
Publication No. US2004003488A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 44229  
LENGTH: 355  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: 700876446\_FLI.pep  
US-10-425-114-44229  
Query Match 45.6%; Score 47; DB 12; Length 355;  
Best Local Similarity 45.5%; Pred. No. 96;  
Matches 10; Conservative 4; Mismatches 6; Indels 2; Gaps 1;  
QY 1 QRLKYAQEELSN--EVLPPPK 20  
DB 202 EMLPYLNEKLKRNVEVPPPK 223  
RESULT 18  
US-10-369-493-1534  
Sequence 1534, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 1534  
LENGTH: 592  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-1534  
Query Match 45.6%; Score 47; DB 15; Length 592;  
Best Local Similarity 50.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 7 QEELSNEVLPPPK 20  
DB 369 EEEMENKAPPKK 382  
RESULT 19  
US-10-369-493-5013  
Sequence 5013, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragoras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 5789:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 444 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...444  
SEQUENCE DESCRIPTION: SEQ ID NO: 5789:  
US-10-335-977-5789  
Query Match 46.6%; Score 48; DB 12; Length 444;  
Best Local Similarity 75.0%; Pred. No. 87;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 9 ELSNEVLPPPK 20  
DB 416 EADNEELPPPK 427  
RESULT 16  
US-10-424-599-234250  
Sequence 234250, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 234250  
LENGTH: 259  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_53552C.1.pep  
US-10-424-599-234250  
Query Match 45.6%; Score 47; DB 12; Length 259;  
Best Local Similarity 45.5%; Pred. No. 88;  
Matches 10; Conservative 4; Mismatches 6; Indels 2; Gaps 1;  
QY 1 QRLKYAQEELSN--EVLPPPK 20  
DB 106 EMLPYLNEKLKRNVEVPPPK 127  
RESULT 17  
US-10-425-114-44229  
Sequence 44229, Application US/10425114  
Publication No. US2004003488A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 44229  
LENGTH: 355  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: 700876446\_FLI.pep  
US-10-425-114-44229  
Query Match 45.6%; Score 47; DB 12; Length 355;  
Best Local Similarity 45.5%; Pred. No. 96;  
Matches 10; Conservative 4; Mismatches 6; Indels 2; Gaps 1;  
QY 1 QRLKYAQEELSN--EVLPPPK 20  
DB 202 EMLPYLNEKLKRNVEVPPPK 223  
RESULT 18  
US-10-369-493-1534  
Sequence 1534, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 1534  
LENGTH: 592  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-1534  
Query Match 45.6%; Score 47; DB 15; Length 592;  
Best Local Similarity 50.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 7 QEELSNEVLPPPK 20  
DB 369 EEEMENKAPPKK 382  
RESULT 19  
US-10-369-493-5013  
Sequence 5013, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5013
; LENGTH: 6642
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5013

Query Match 45.6%; Score 47; DB 15; Length 6642;
Best Local Similarity 45.0%; Pred. No. 2.3e+03;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 QRLKYAQEELNEVLPPPRK 20
DB 1353 RRVSAEEELPKEVIDSRK 1372

RESULT 20
US-09-171-432A-40
; Sequence 40, Application US/09171432A
; Publication No. US20030187184A1
; GENERAL INFORMATION:
; APPLICANT: Fields, Howard A.
; APPLICANT: Khuyakov, Yuri E.
; TITLE OF INVENTION: Antigenically Reactive Regions of the
; TITLE OF INVENTION: Hepatitis A Virus Polyprotein
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick Stockton LLP
; STREET: 3424 Peachtree Road, N.E.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30326
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/171,432A
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Greene, Jamie L.
; REGISTRATION NUMBER: 32,467
; REFERENCE/DOCKET NUMBER: 03063-0231US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 949-2400
; TELEFAX: (404) 949-2499
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..20
; OTHER INFORMATION: /label=YK-1316
US-09-171-432A-40

Query Match 43.7%; Score 45; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QRLKYAQEE 9
DB 12 QRLKYAQEE 20

RESULT 21
US-10-424-599-234251
; Sequence 234251, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234251
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(212)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_53553C.1.pcp
US-10-424-599-234251

Query Match 43.7%; Score 45; DB 12; Length 212;
Best Local Similarity 40.9%; Pred. No. 1.1e+02;
Matches 9; Conservative 5; Mismatches 6; Indels 2; Gaps 1;

QY 1 QRLKYAQEELSN--EVLPPPRK 20
DB 132 EMVSYLNEXLKRNVVPPPKK 153

RESULT 22
US-09-864-761-37656
; Sequence 37656, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30



PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
 SEQ ID NO 37656
 LENGTH: 263
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC002346.1
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
 OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 6
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
 OTHER INFORMATION: EST HUMAN HIT: AUI38405.1, EVALUE 7.00e-90
 OTHER INFORMATION: SWISSPROT HIT: P10269, EVALUE 1.00e-16
 US-09-864-761-37656

 Query Match 43.7%; Score 45; DB 9; Length 263;
 Best Local Similarity 46.7%; Pred. No. 1.4e+02;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

 QY 5 YAEQDLYNHVPPKPR 19
 DB 174 YTEQDLYNHVPPKPR 188

 RESULT 23
 US-10-424-599-234246
 Sequence 234246, Application US/10424599
 Publication No. US20040031072A1
 GENERAL INFORMATION:
 APPLICANT: La Rosa Thomas J
 APPLICANT: Kovalic David K
 APPLICANT: Zhou Yihua
 APPLICANT: Cao Yongwei
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53223)B
 CURRENT APPLICATION NUMBER: US/10/424,599
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 285684
 SEQ ID NO 234246
 LENGTH: 322
 TYPE: PRT
 ORGANISM: Glycine max
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT\_MRT3847\_53549C.1.pep
 US-10-424-599-234246

 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
 SEQ ID NO 37656
 LENGTH: 263
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC002346.1
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
 OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 6
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
 OTHER INFORMATION: EST HUMAN HIT: AUI38405.1, EVALUE 7.00e-90
 OTHER INFORMATION: SWISSPROT HIT: P10269, EVALUE 1.00e-16
 US-09-864-761-37656

 Query Match 43.7%; Score 45; DB 9; Length 263;
 Best Local Similarity 46.7%; Pred. No. 1.4e+02;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

 QY 5 YAEQDLYNHVPPKPR 19
 DB 174 YTEQDLYNHVPPKPR 188

 RESULT 23
 US-10-424-599-234246
 Sequence 234246, Application US/10424599
 Publication No. US20040031072A1
 GENERAL INFORMATION:
 APPLICANT: La Rosa Thomas J
 APPLICANT: Kovalic David K
 APPLICANT: Zhou Yihua
 APPLICANT: Cao Yongwei
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53223)B
 CURRENT APPLICATION NUMBER: US/10/424,599
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 285684
 SEQ ID NO 234246
 LENGTH: 322
 TYPE: PRT
 ORGANISM: Glycine max
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT\_MRT3847\_53549C.1.pep
 US-10-424-599-234246

 Query Match 43.7%; Score 45; DB 12; Length 322;
 Best Local Similarity 40.9%; Pred. No. 1.7e+02;
 Matches 9; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

 QY 1 QRLKYAOEELS--EVLPPPK 20
 DB 130 EWSYLNKLEKRNVEVPPPK 201

 RESULT 24
 US-10-389-566-1970
 Sequence 1970, Application US/10389566
 Publication No. US20040025202A1
 GENERAL INFORMATION:
 APPLICANT: Monsanto Technology, LLC
 APPLICANT: Laurie, Cathy C
 TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
 FILE REFERENCE: 38-77(52900)D
 CURRENT APPLICATION NUMBER: US/10/389,566
 CURRENT FILING DATE: 2003-03-31
 PRIOR APPLICATION NUMBER: US 60/365,301
 PRIOR FILING DATE: 2002-03-15
 PRIOR APPLICATION NUMBER: US 60/391,786
 PRIOR FILING DATE: 2002-06-25
 PRIOR APPLICATION NUMBER: US 60/392,018
 PRIOR FILING DATE: 2002-06-26
 NUMBER OF SEQ ID NOS: 2459
 SOFTWARE: PatentIn version 3.2
 SEQ ID NO 1970
 LENGTH: 756
 TYPE: PRT
 ORGANISM: Pisum sativum
 US-10-389-566-1970

 Query Match 43.7%; Score 45; DB 16; Length 756;
 Best Local Similarity 60.0%; Pred. No. 4.3e+02;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

 QY 2 RLKYAOEELSNEVL 16
 DB 64 RLQFAKPNMNEVL 78

 RESULT 25
 US-10-369-493-3980
 Sequence 3980, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 3980
 LENGTH: 1168
 TYPE: PRT
 ORGANISM: Neurospora crassa
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (1)..(1168)
 OTHER INFORMATION: unsure at all Xaa locations
 US-10-369-493-3980

 Query Match 43.7%; Score 45; DB 15; Length 1168;

Best Local Similarity 43.8%; Pred. No. 6.9e+02;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
QY 5 YAOEELSNEVLPPPK 20  
Db 246 YAEEDSDELLPKPKR 261  
RESULT 26  
US-09-909-567B-47  
; Sequence 47, Application US/09909567B  
; Publication No. US2003002257A1  
; GENERAL INFORMATION:  
; APPLICANT: Macina, Roberto A.  
; APPLICANT: Nair, Manoj  
; APPLICANT: Chen, Seiyu  
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes  
; FILE REFERENCE: DEX-0214  
; CURRENT APPLICATION NUMBER: US/09/909,567B  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: 60/219,834  
; PRIOR FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 47  
; LENGTH: 1596  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-909-567B-47  
Query Match 43.7%; Score 45; DB 10; Length 1596;  
Best Local Similarity 43.8%; Pred. No. 9.7e+02;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 3 LKAYAEELSNEVLPPPK 18  
Db 799 LNYEEKDSEDQVLPPK 814  
RESULT 27  
US-10-425-114-57289  
; Sequence 57289, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 57289  
; LENGTH: 235  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mexicana  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMROTEOSINTE091H01\_FLI.pep  
US-10-425-114-57289

Query Match 43.2%; Score 44.5; DB 12; Length 235;  
Best Local Similarity 55.0%; Pred. No. 1.4e+02;  
Matches 11; Conservative 0; Mismatches 6; Indels 3; Gaps 1;  
QY 1 QRLKYAEELSNEVLPPPK 20  
Db 217 QRLSYAQAPFSG---PPPK 233

RESULT 28  
US-10-425-114-72504  
; Sequence 72504, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 72504  
; LENGTH: 459  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMFLB73256H07\_FLI.pep  
US-10-425-114-72504

Query Match 43.2%; Score 44.5; DB 12; Length 459;  
Best Local Similarity 55.0%; Pred. No. 3e+02;  
Matches 11; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

QY 1 QRLKYAEELSNEVLPPPK 20  
Db 441 QRLSYAQAPFSG---PPPK 457

RESULT 29  
US-10-424-599-239647  
; Sequence 239647, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 239647  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(221)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT MRT3847\_58426C.1.pep  
US-10-424-599-239647

Query Match 42.7%; Score 44; DB 12; Length 221;  
Best Local Similarity 45.0%; Pred. No. 1.6e+02;  
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 QRLKYAEELSNEVLPPPK 20  
Db 70 RKTQKQERSNEEEPPQE 89

RESULT 30  
US-10-369-493-22337  
; Sequence 22337, Application US/10369493

; Publication No. US20030333675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 39-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 22337  
; LENGTH: 303  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-22337

Query Match 42.7%; Score 44; DB 15; Length 303;  
Best Local Similarity 56.2%; Pred. No. 2.2e+02;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QRLKYAQEELSNEVLP 16  
Db 183 RHLKYGPEKNERLP 198  
: ||| : |||  
: ||| : |||

Search completed: May 11, 2004, 13:44:50  
Job time : 35.25 secs



99 41 40.6 108 2 AAW69486 Aaw69486 Hepatitis  
100 41 40.6 113 2 AAY48558 Aay48558 Human bre

ALIGNMENTS

RESULT 1  
AAW42925  
ID AAW42925 standard; peptide; 20 AA.  
XX  
AC AAW42925;  
XX  
DT 28-APR-1998 (first entry)  
XX  
DE Immunogenic Hepatitis A virus peptide YK-1318.  
XX  
KW Immunogenic peptide; immunogenic epitope; P2A protein; immune response;  
KW antibody.  
XX  
OS Synthetic.  
OS Hepatitis A virus.  
XX  
PN W09740147-A1.  
XX  
PD 30-OCT-1997.  
XX  
PF 18-APR-1997; 97WO-US006891.  
XX  
PR 19-APR-1996; 96US-0015644P.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI; 1997-535831/49.  
XX  
PT Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
PT response to HAV in a mammal or to detect the presence of antibodies  
PT against HAV in a mammal.  
XX  
PS Claim 18; Page 112; 140pp; English.  
XX  
CC Peptides AAW42922-30 are immunogenic peptides corresponding to  
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
CC substantially similar to a portion of the amino acid sequence of the P2A  
CC protein of HAV corresponding to amino acids 792-980. The present peptide  
CC is derived from amino acids 823-842, and has a reactivity of 31.3% with  
CC acute sera. Compositions containing the peptides can be used to induce an  
CC immune response to HAV in a mammal. The peptides can also be used to  
CC detect the presence of antibodies against HAV in mammalian serum. The  
CC peptides can also be used to make an antibody against HAV by  
CC administering the peptide to a mammal  
XX  
SQ Sequence 20 AA;

Query Match 100.0%; Score 101; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VLPPPRKMKGLFSQAKISLF 20  
DB 1 VLPPPRKMKGLFSQAKISLF 20

RESULT 2  
AAB69442  
ID AAB69442 standard; peptide; 21 AA.  
XX  
AC AAB69442;  
XX  
DT 20-APR-2001 (first entry)  
XX

DE Synthetic HAV P2A peptide, SEQ ID NO: 42.  
XX  
KW Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
KW antigen; major structural capsid polypeptide; HAV antibody detection.  
XX  
OS Hepatitis A virus.  
OS Synthetic.  
XX  
RN W0200105824-A2.  
XX  
PD 25-JAN-2001.  
XX  
PF 14-JUL-2000; 2000WO-US019267.  
XX  
PR 15-JUL-1999; 99US-0144412P.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Fields HA, Khudyakov YE;  
XX  
DR WPI; 2001-112681/12.  
XX  
PT Synthetic peptides used as antigen sources for enzyme immunoassays  
PT detecting anti-hepatitis A virus and as vaccines.  
XX  
PS Claim 13; Page 95; 130pp; English.

The present sequence is one of a number of synthetic peptides which are  
immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
comprise antigenic epitopes of the major structural capsid polypeptides  
or non-structural polypeptides of HAV with one or more glutamine  
molecules at the carboxy end of the peptide. The peptides are used to  
detect the presence of antibodies against HAV in mammalian serum, to  
detect the presence of HAV in a human or animal through the binding of  
the peptide to an antibody, to detect acute phase infection by detecting  
IgM antibodies in mammalian serum and detecting convalescence in a  
mammal. The peptides are used to detect or quantify HAV antibodies in  
samples in clinical or research-based assays using immunoblotting,  
fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
tracking of radioactive or bioluminescent markers, chromatography or  
electrophoresis. The peptides are used to induce an immune response to  
HAV when administered to a human or animal. Glutamine at the carboxy end  
of the peptides enhances the IgM antibody reactivity

Sequence 21 AA;  
Query Match 100.0%; Score 101; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 5.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
DB 1 VLPPPRKMKGLFSQAKISLF 20

RESULT 3  
AAP50230  
ID AAP50230 standard; protein; 366 AA.  
XX  
AC AAP50230;  
XX  
DT 28-NOV-1991 (first entry)  
XX  
DE Sequence of hepatitis A virus (HAV) surface protein (VP-1).  
XX  
KW Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
KW diagnostic assay.  
XX  
OS Hepatitis A virus.  
XX  
PN EP138704-A.  
XX  
PD 24-APR-1985.

XX 09-OCT-1984; 84EP-00402025.  
 XX 14-OCT-1983; 83US-00541836.  
 PR 02-MAR-1984; 84US-00585942.  
 XX (MERI ) MERCK & CO INC.  
 XX Hughes JV, Scolnick EM, Tomassini JE;  
 XX WPI; 1985-100818/17.  
 DR N-PSDB; AAN50274.  
 XX New hepatitis A virus surface protein - useful for binding to  
 PT neutralising antibodies to the virus.  
 XX Claim 21; Page 46-48; 49pp; English.  
 CC VP1 is isolated by solubilisation of the intact virus in an aq. anionic  
 CC surfactant and a reducing agent. The viral proteins are sepd. and the  
 CC protein of molecular wt. 33000 daltons is sepd  
 XX Sequence 366 AA;  
 SQ  
 Query Match 100.0%; Score 101; DB 1; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 1e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLPPEKMKGLFSQAKISLF 20  
 DB 332 VLPPEKMKGLFSQAKISLF 351  
 RESULT 4  
 AAP50287  
 ID AAP50287 standard; protein; 854 AA.  
 XX AAP50287;  
 AC AAP50287;  
 XX 25-MAR-2003 (revised)  
 DT 30-NOV-1991 (first entry)  
 DE Hepatitis A virus (HAV) peptide corresponding to the capsid protein  
 DE region of poliovirus RNA.  
 XX Hepatitis A virus assay; antigen; antibody.  
 XX Hepatitis A virus.  
 OS WO8501517-A.  
 PN 11-APR-1985.  
 XX 27-SEP-1984; 84WO-US0001552.  
 XX 30-SEP-1983; 83US-00537911.  
 XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX Ticehurst JR, Baltimore D, Feinstein SM, Purcell RH;  
 PI Racanello VR;  
 XX WPI; 1985-098846/16.  
 DR N-PSDB; AAN50330.  
 XX New hepatitis A virus CDNA - useful in assays for the virus and for  
 PT prodn. of the viral antigen and antibodies to it.  
 XX Example; Fig 7; 60pp; English.  
 PS The inventors claim HAV cDNA and a method for producing it, whereby large  
 CC ants. can be obtd. economically. The cDNA is useful in the assay for  
 CC detection of HAV quickly and easily and with high sensitivity and

CC specificity. The HAV cDNA is also used in the prodn. of HAV antigen or  
 CC antibodies to it. The antibodies may be monoclonal. (Updated on 25-MAR-  
 CC 2003 to correct PA field.)  
 XX Sequence 854 AA;  
 SQ  
 Query Match 100.0%; Score 101; DB 1; Length 854;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLPPEKMKGLFSQAKISLF 20  
 DB 823 VLPPEKMKGLFSQAKISLF 842  
 RESULT 5  
 AAP50116  
 ID AAP50116 standard; protein; 993 AA.  
 XX AAP50116;  
 AC AAP50116;  
 XX 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 30-SEP-1991 (first entry)  
 XX Sequence of Hepatitis A virus (HAV) immunogenic peptides VP-1, VP-2, VP-3  
 DE and VP-4.  
 DE Antigenic protein; immunogen; vaccine.  
 KW Hepatitis A virus; (strain CR326).  
 OS EPI54587-A.  
 PN 11-SEP-1985.  
 PD 27-FEB-1985; 85EP-00400369.  
 XX 02-MAR-1984; 84US-00585818.  
 XX (MERI ) MERCK & CO INC.  
 PA Linemeyer DL, Menke JG, Rueben RG, Mitra SW;  
 PI WPI; 1985-224964/37.  
 DR N-PSDB; AAN50139.  
 XX New nucleotide sequences coding for hepatitis A virus antigens - useful  
 PT for eliciting normal immune response and in vaccines for protecting  
 PT against the virus.  
 XX Example; Page 11-17; 32pp; English.  
 PS Within the sequence in AAN50139 is encoded the information necessary to  
 CC make the antigenic proteins of HAV. The sequences encoding for the  
 CC structural proteins begin at base 403. The key sub-unit sequences within  
 CC VP-1, designated Sequences I, II, III, IV, and V, start, respectively at  
 CC 1882, 1963, 1999, 2146, 2347. Other nucleotide sequences which are  
 CC valuable as encoding antigenic proteins are the sequences from base 1749  
 CC to base 2722; from base 1487 to base 2980 and from base 1644 to base  
 CC 2722. The sequence from base 1749 to base 2722 is esp. valuable as a  
 CC vector for producing antigen protein. Sequences II-V are claimed. X in  
 CC AAP50116 denotes the translation of a stop codon. (Updated on 25-MAR-2003  
 CC to correct PI field.) (Updated on 24-OCT-2003 to standardise OS field)  
 XX Sequence 993 AA;  
 SQ  
 Query Match 100.0%; Score 101; DB 1; Length 993;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLPPEKMKGLFSQAKISLF 20  
 DB 823 VLPPEKMKGLFSQAKISLF 842

Db 959 VLPPPRKMKGLFSQAKISLF 978

RESULT 6  
AAP50231  
ID AAP50231 standard; protein; 993 AA.  
AC AAP50231;  
XX  
XX 28-NOV-1991 (first entry)  
DT  
XX  
DE Sequence encoded by partial sequence of hepatitis A virus (HAV),  
DE including surface protein (VP-1).  
XX  
XX Hepatitis A virus vaccine; immunisation; monoclonal antibody;  
KW diagnostic assay.  
XX  
XX Hepatitis A virus.  
OS  
XX  
XX Key Location/Qualifiers  
FH 628..993  
FT Protein  
FT /note= "claimed; X denotes translated stop codons and  
FT unspecified triplets"  
XX  
XX EPI38704-A.  
PN  
XX  
XX 24-APR-1985.  
PD  
XX  
XX 09-OCT-1984; 84EP-00402025.  
PF  
XX 14-OCT-1983; 83US-00541836.  
PR  
XX 02-MAR-1984; 84US-00585942.  
PR  
XX (MERI ) MERCK & CO INC.  
PA  
XX  
XX Hughes JV, Scolnick EM, Tomassini JE;  
PI  
XX  
XX WPI; 1985-100818/17.  
DR  
XX N-PSDB; AAN50274.  
DR  
XX New hepatitis A virus surface protein - useful for binding to  
FT neutralising antibodies to the virus.  
PT  
XX  
XX Disclosure; Page 17-23; 49pp; English.  
PS  
XX  
XX VP1 is isolated by solubilisation of the intact virus in an aq. anionic  
CC surfactant and a reducing agent. The viral proteins are sepd. and the  
CC protein of molecular wt. 33000 daltons is sepd  
CC  
XX  
XX Sequence 993 AA;  
SQ

Query Match 100.0%; Score 101; DB 1; Length 993;  
Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
DB 959 VLPPPRKMKGLFSQAKISLF 978

RESULT 7  
AAW95559  
ID AAW95559 standard; protein; 1077 AA.  
XX  
XX  
XX AAW95559;  
AC  
XX  
XX 28-APR-1999 (first entry)  
DT  
XX  
XX A partial hepatitis A virus (HAV) protein.  
DE  
XX  
XX Hepatitis A virus protein; HAV; P2 region;  
KW cell-culture-adapted HAV strain; infection; accelerated growth.  
KW  
XX

OS Hepatitis A virus.  
XX  
XX US5849562-A.  
PN  
XX  
XX 15-DEC-1998.  
PD  
XX  
XX 06-JUN-1995; 95US-00468926.  
PF  
XX  
XX 30-SEP-1983; 83US-00537911.  
PR  
XX 27-SEP-1984; 84US-0054942.  
PR  
XX 06-OCT-1988; 88US-00256135.  
PR  
XX 06-NOV-1991; 91US-00788262.  
PR  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA  
XX  
XX Emerson SU, Purcell RH;  
PI  
XX  
XX WPI; 1999-094412/08.  
DR  
XX N-PSDB; AAX01006.  
DR  
XX Chimeric hepatitis A virus strains - with P2 region from cell-culture-  
FT adapted strain in wild-type genome.  
FT  
XX  
XX Disclosure; Fig 7A-L; 36pp; English.  
PS  
XX  
XX The present sequence represents a partial hepatitis A virus (HAV)  
CC protein. The specification describes a DNA construct consisting of a wild  
CC -type HAV genome in which the P2 region is replaced by the P2 region from  
CC a cell-culture-adapted HAV strain. The construct is used to demonstrate  
CC that mutations in the P2 region of a cell-culture-adapted HAV strain are  
CC sufficient for establishment of infection and accelerated growth in cell  
CC culture  
XX  
XX Sequence 1077 AA;  
SQ

Query Match 100.0%; Score 101; DB 2; Length 1077;  
Best Local Similarity 100.0%; Pred. No. 3e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
DB 1046 VLPPPRKMKGLFSQAKISLF 1065

RESULT 8  
AAR32426  
ID AAR32426 standard; protein; 1091 AA.  
XX  
XX AAR32426;  
AC  
XX  
XX 27-AUG-2003 (revised)  
DT  
XX 25-MAR-2003 (revised)  
DT  
XX 17-DEC-2001 (revised)  
DT  
XX 10-JUN-1993 (first entry)  
DT  
XX Translated from 5' region of Hepatitis A Virus genomic clone.  
DE  
XX HAV HM-175; chronic liver disease; picornavirus.  
KW  
XX  
XX Hepatitis A virus.  
OS  
XX  
XX Key Location/Qualifiers  
FH 1..711  
FT Region  
FT /note= "X's correspond to nonsense codons, i.e. this  
FT region is not an ORF"  
FT 238..1091  
FT /label= ORF  
FT /note= "second putative initiation codon at position 240"  
XX  
XX USN7788262-N.  
PN  
XX  
XX 15-DEC-1992.  
PD  
XX

PF 06-NOV-1991; 91US-00788262.  
 XX 30-SEP-1983; 88US-00536911.  
 PR 27-SEP-1984; 84US-00654942.  
 PR 06-OCT-1988; 88US-00256135.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 XX  
 XX Ticehurst JR, Baltimore D, Feinstein SM, Purcell RH;  
 PI Racanelli VR, Baroudy BM, Emerson SU;  
 XX  
 XX WPI; 1993-067429/08.  
 DR N-PSDB; AAQ36934.  
 DR  
 XX Hepatitis A virus cDNA prodn. - for diagnostic use and for prodn. of  
 PT antigen and antibodies.  
 PT  
 XX Disclosure; Fig 7; 65pp; English.  
 PS  
 XX HAV virion RNA was extracted from the livers of marmosets which had been  
 CC inoculated with HAV (the HAV had previously been passaged twice in  
 CC marmosets). The RNA was used to prepare ds cDNA clones by standard  
 CC methods. Clones contg. inserts which hybridised to RNA from HAV-infected  
 CC African Green Monkey Kidney cells were selected for further analysis. A  
 CC 7.4kb restriction map (about 99% of the HAV genome) was constructed from  
 CC 5' overlapping inserts. The sequence of the first 3.3kb (approx.) from the  
 CC 5'-terminus was determined. An amino acid sequence was decoded from the  
 CC entire clone and an open reading frame was identified starting at  
 CC position 238. A comparison of the predicted HAV amino acid sequences with  
 CC the known capsid protein sequences of other picornaviruses (poliovirus,  
 CC foot and mouth disease virus and encephalomyelitis virus) revealed areas  
 CC of local homology. (Note: Revised entry submitted to correct the patent  
 CC number format of US Government-owned NTIS applications to prevent clashes  
 CC with ongoing US granted patent numbers. For further information please  
 CC visit the Derwent web site at [www.derwent.com/dwpi/updates/ntis.us.html](http://www.derwent.com/dwpi/updates/ntis.us.html).)  
 CC (Updated on 25-MAR-2003 to correct PI field.) (Updated on 27-AUG-2003 to  
 CC correct OS field.)  
 XX  
 XX Sequence 1091 AA;  
 SQ  
 Query Match 100.0%; Score 101; DB 2; Length 1091;  
 Best Local Similarity 100.0%; Pred. No. 3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLPPPRKMKGLFSQAKISLF 20  
 DB 1060 VLPPPRKMKGLFSQAKISLF 1079  
 RESULT 9  
 AAR05697  
 ID AAR05697 standard; protein; 2227 AA.  
 XX  
 AC AAR05697;  
 XX  
 XX 24-OCT-2003 (revised)  
 DT 27-AUG-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 15-AUG-1990 (first entry)  
 XX  
 XX Attenuated hepatitis A virus.  
 XX  
 XX Hepatitis A virus; vaccine; attenuated.  
 KW  
 XX Hepatitis A virus; strain HM-175.  
 OS  
 XX Key Location/Qualifiers  
 FH Region 1..23  
 FT /label= VP4 = 1A  
 FT Region 24..245  
 FT /label= VP2 = 1B  
 FT Region 246..491  
 FT /label= VP3 = 1C

FT Region 492..791  
 FT /label= VP1 = 1D  
 FT Region 792..980  
 FT /label= 2A  
 FT Region 981..1087  
 FT /label= 2B  
 FT Region 1088..1422  
 FT /label= 2C  
 FT Region 1423..1496  
 FT /label= 3A  
 FT Region 1497..1519  
 FT /label= 3B = VPg  
 FT Region 1520..1738  
 FT /label= 3C  
 FT Region 1739..2227  
 FT /label= 3D  
 XX  
 PN US4894228-A.  
 XX  
 XX 16-JAN-1990.  
 PD  
 XX 12-JUL-1988; 88US-00217824.  
 PF  
 XX 19-SEP-1984; 84US-00652067.  
 PR  
 XX 09-SEP-1986; 86US-00905146.  
 PR  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
 PA  
 XX Purcell RH, Ticehurst JR, Cohen JL, Emerson SU, Feinstein SM;  
 PI Daemer RJ, Gust ID;  
 PI  
 XX WPI; 1990-075557/10.  
 DR N-PSDB; AAQ03512.  
 DR  
 XX Vaccine against hepatitis A virus infection - comprises novel attenuated  
 PT hepatitis A virus strain.  
 PT  
 XX Claim 1; Fig 1; 18pp; English.  
 PS  
 XX The attenuated HAV is useful for inducing protective immunity against  
 CC HAV. This strain (Pass 35) differs from the wild type HAV HM-175 by  
 CC several nucleotide changes distributed throughout the genome, is  
 CC attenuated for chimpanzees, elicits serum neutralising antibodies, and is  
 CC suitable for use as an HAV vaccine. It is noted that not all the changes  
 CC are necessary for attenuation and use as a vaccine. (Updated on 25-MAR-  
 CC 2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PI field.)  
 CC (Updated on 27-AUG-2003 to correct OS field.) (Updated on 24-OCT-2003 to  
 CC standardise OS field)  
 XX  
 XX Sequence 2227 AA;  
 SQ  
 Query Match 100.0%; Score 101; DB 2; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLPPPRKMKGLFSQAKISLF 20  
 DB 823 VLPPPRKMKGLFSQAKISLF 842  
 RESULT 10  
 AAW34074  
 ID AAW34074 standard; protein; 2227 AA.  
 XX  
 AC AAW34074;  
 XX  
 XX 17-OCT-2003 (revised)  
 DT 27-OCT-1998 (first entry)  
 DT  
 XX Hepatitis A virus HM-175 protein sequence.  
 DE  
 XX HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus; infection;  
 KW HAV vaccine.





XX 05-SEP-2000.  
XX  
XX  
XX PF 07-JUN-1995; 95US-00475886.  
XX  
XX PR 18-SEP-1992; 92US-00947338.  
XX PR 17-SEP-1993; 93WO-US008610.  
XX PR 17-APR-1995; 95US-00397232.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX PA  
XX PI D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
XX DR WPI; 2000-586464/55.  
XX DR N-PSDB; AAA75476.  
XX  
XX PT Novel live hepatitis A virus adapted to growth in human fibroblast cell  
XX PT line useful as vaccine for protecting humans against hepatitis A virus  
XX PT infection, has modified genome compared to wild type.  
XX PS  
XX PS Disclosure; Fig 6A-K; 72pp; English.  
XX  
XX CC The present sequence is derived from a wild type hepatitis A virus (HAV)  
XX CC strain HM-174. The sequence is modified to produce HAV which are adapted  
XX CC to growth in the human fibroblast-like cell line MRC-5. The HAV is able  
XX CC to propagate in MRC-5 cells and retain appropriate attenuation. It is  
XX CC useful as a live vaccine for prophylaxis of hepatitis A in humans and  
XX CC other primates  
XX  
XX SQ Sequence 2227 AA;  
Query Match 100.0%; Score 101; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 6.1e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VLPPPRKMKGLFSQAKISLF 20  
DB 823 VLPPPRKMKGLFSQAKISLF 842  
RESULT 13  
AAB18608  
ID AAB18608 standard; protein; 2227 AA.  
XX  
XX AC AAB18608;  
XX  
XX DT 15-JAN-2001 (first entry)  
XX  
XX DE Amino acid sequence of passage 35 Hepatitis A virus called P-35.  
XX KW HAV; strain HM-174; MRC-5 cell; live vaccine; hepatitis A infection;  
XX KW P-35 virus.  
XX  
XX OS Hepatitis A virus.  
XX  
XX PN US6113912-A.  
XX  
XX PD 05-SEP-2000.  
XX  
XX PF 07-JUN-1995; 95US-00475886.  
XX  
XX PR 18-SEP-1992; 92US-00947338.  
XX PR 17-SEP-1993; 93WO-US008610.  
XX PR 17-APR-1995; 95US-00397232.  
XX  
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
XX PI D'hondt E, Purcell RH, Emerson SU, Funkhouser AW;  
XX XX WPI; 2000-586464/55.  
XX XX DR N-PSDB; AAA75477.  
XX  
XX PT Novel live hepatitis A virus adapted to growth in human fibroblast cell

PT line useful as vaccine for protecting humans against hepatitis A virus  
PT infection, has modified genome compared to wild type.  
XX  
XX PS Disclosure; Col 67-78; 72pp; English.  
XX  
XX CC The present sequence is derived from passage 35 of a wild type hepatitis  
XX CC A virus (HAV) strain HM-174. The resulting virus is designated P-35  
XX CC virus. The sequence is modified to produce HAV which are adapted to  
XX CC growth in the human fibroblast-like cell line MRC-5. The HAV is able to  
XX CC propagate in MRC-5 cells and retain appropriate attenuation. It is useful  
XX CC as a live vaccine for prophylaxis of hepatitis A in humans and other  
XX CC primates  
XX  
XX SQ Sequence 2227 AA;  
Query Match 100.0%; Score 101; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 6.1e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VLPPPRKMKGLFSQAKISLF 20  
DB 823 VLPPPRKMKGLFSQAKISLF 842  
RESULT 14  
AAE19899  
ID AAE19899 standard; protein; 2227 AA.  
XX  
XX AC AAE19899;  
XX  
XX DT 18-JUN-2002 (first entry)  
XX  
XX DE Hepatitis A virus (HAV) protein.  
XX KW Hepatitis A virus; HAV; infection; virucide; fungicide; antibacterial;  
XX KW cytostatic; immunostimulant; vaccine; ribavirin; immune response; cancer.  
XX OS Hepatitis A virus.  
XX  
XX PN WO200213855-A2.  
XX  
XX PD 21-FEB-2002.  
XX  
XX PF 15-AUG-2001; 2001WO-IB001808.  
XX  
XX PR 17-AUG-2000; 2000US-0225767P.  
XX PR 29-AUG-2000; 2000US-0229175P.  
XX PR 03-NOV-2000; 2000US-00705547.  
XX  
XX PA (TRIP-) TRIPEP AB.  
XX  
XX PI Sallberg M, Hultgren C;  
XX  
XX DR WPI; 2002-241837/29.  
XX DR N-PSDB; AAD31766.  
XX  
XX PT Vaccine compositions for treating and preventing disease, preferably  
XX PT hepatitis C virus infection, comprises ribavirin and antigen that has  
XX PT epitope present in hepatitis C virus.  
XX  
XX PS Claim 11; Page 92-87; 120pp; English.  
XX  
XX CC The invention relates to a composition comprising ribavirin and an  
XX CC antigen preferably non structural 3 protein (NS3)/4A fragment of  
XX CC hepatitis C virus (HCV) genome or a peptide or nucleic acid of HCV  
XX CC sequence. The composition is useful for enhancing an immune response to a  
XX CC hepatitis C antigen in humans, domestic, sport or pet species and as  
XX CC vaccines for treating and preventing HCV infections. The composition is  
XX CC also useful for treating viral, bacterial, fungal diseases and cancer.  
XX CC The present sequence is hepatitis A virus (HAV) protein  
XX  
XX SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 5; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
 |||||  
 DB 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 15  
 ABG31729  
 ID ABG31729 standard; protein; 2227 AA.  
 XX AC ABG31729;  
 XX DT 29-AUG-2003 (revised)  
 XX DT 29-NOV-2002 (first entry)  
 XX DE Attenuated Hepatitis A virus (HAV) 4380 polypeptide.  
 XX KW Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast; virucide;  
 XX KW HAV 4380.  
 XX OS Hepatitis A virus; strain HM-175.  
 XX PN US6423318-B1.  
 XX PD 23-JUL-2002.  
 XX PF 31-AUG-2000; 2000US-00653499.  
 XX PR 17-SEP-1993; 93WO-US008610.  
 XX PR 17-APR-1995; 95US-00397232.  
 XX PR 07-JUN-1995; 95US-00475886.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX PI Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
 XX WPI; 2002-680946/73.  
 XX N-PSDB; ABS52789.  
 XX New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth  
 in MRC-5 cells, useful for preparing a vaccine against HAV infection.  
 Disclosure; Col 93-104; 71pp; English.

QY 1 VLPPPRKMKGLFSQAKISLF 20  
 |||||  
 DB 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 16  
 ABG31727  
 ID ABG31727 standard; protein; 2227 AA.  
 XX AC ABG31727;  
 XX DT 29-AUG-2003 (revised)

DT 29-NOV-2002 (first entry)  
 XX DE Wild-type Hepatitis A virus (HAV) strain HM-175 polypeptide.  
 XX KW Hepatitis A virus; HAV; HM-175; MRC-5 cell; human fibroblast; virucide.  
 XX OS Hepatitis A virus; strain HM-175.  
 XX PN US6423318-B1.  
 XX PD 23-JUL-2002.  
 XX PF 31-AUG-2000; 2000US-00653499.  
 XX PR 17-SEP-1993; 93WO-US008610.  
 XX PR 17-APR-1995; 95US-00397232.  
 XX PR 07-JUN-1995; 95US-00475886.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX PI Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
 XX WPI; 2002-680946/73.  
 XX N-PSDB; ABS52789.  
 XX New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth  
 in MRC-5 cells, useful for preparing a vaccine against HAV infection.  
 Disclosure; Fig 6; 71pp; English.

QY 1 VLPPPRKMKGLFSQAKISLF 20  
 |||||  
 DB 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 17  
 ABG31728  
 ID ABG31728 standard; protein; 2227 AA.  
 XX AC ABG31728;  
 XX DT 29-NOV-2002 (first entry)  
 XX DE Hepatitis A virus mutant strain HM-175/7 (pHAV/7) polypeptide.  
 XX KW Hepatitis A virus; HAV; HM-175; HM-175/7; MRC-5 cell; human fibroblast;  
 XX KW virucide; mutant; pHAV/7; mutein.  
 XX OS Hepatitis A virus; strain HM-175.  
 XX OS Synthetic.  
 XX Key Location/Qualifiers  
 XX FT Misc-difference 764  
 XX FT Misc-difference 821 /note= "Wild-type Glu substituted by Val"  
 XX FT Misc-difference 821 /note= "Wild-type Asn substituted by Ser"  
 XX FT Misc-difference 963 /label= Wild-type Lys substituted by Arg  
 XX FT Misc-difference 1052

FT Misc-difference 1062 /note= "Wild-type Ala substituted by Val"

FT Misc-difference 1118 /note= "Wild-type Gly substituted by Ala"

FT Misc-difference 1151 /note= "Wild-type Lys substituted by Met"

FT Misc-difference 1163 /note= "Wild-type Glu substituted by Lys"

FT Misc-difference 1277 /note= "Wild-type Phe substituted by Ser"

FT Misc-difference 1500 /note= "Wild-type Val substituted by Ile"

FT Misc-difference 1805 /note= "Wild-type His substituted by Tyr"

FT Misc-difference 1930 /note= "Wild-type Asp substituted by Asn"

FT Misc-difference 1930 /note= "Wild-type Ser substituted by Thr"

XX US6423318-B1.

XX 23-JUL-2002.

XX 31-AUG-2000; 2000US-00653499.

PR 17-SEP-1993; 93WO-US008610.

PR 17-APR-1995; 95US-00397232.

PR 07-JUN-1995; 95US-00475886.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;

XX WPI; 2002-680946/73.

DR N-PSDB; ABS52788.

XX New polynucleotide encoding a hepatitis A virus (HAV) adapted to growth in MRC-5 cells, useful for preparing a vaccine against HAV infection.

PT in MRC-5 cells, useful for preparing a vaccine against HAV infection.

XX Example 3; Col 67-78; 7lpp; English.

XX The invention relates to a polynucleotide which encodes a hepatitis A virus (HAV) adapted to growth in MRC-5 cells (human fibroblast-like cell line). The polynucleotide is useful for preparing a vaccine against hepatitis A virus infection. This sequence represents a hepatitis A virus mutant strain HM-175/7 (pHAV/7) polypeptide

XX SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 5; Length 2227;

Best Local Similarity 100.0%; Pred. No. 6.1e-07;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGLFSQAKISLF 20

Db 823 VLPPrKMKGLFSQAKISLF 842

RESULT 18

ABU08640

ID ABU08640 standard; protein; 2227 AA.

XX AC ABU08640;

XX 23-OCT-2003 (revised)

DT 03-JUN-2003 (first entry)

XX Attenuated (pass35) hepatitis A virus strain HM-175.

XX Hepatitis A virus; HAV; virucide; hepatotropic; antiinflammatory; vaccine; MRC-5 cell; hepatitis infection; ds; attenuated virus.

XX Hepatitis A virus; strain HM-175.

XX US2002176869-A1.

PN 28-NOV-2002.

XX 29-APR-2002; 2002US-00135988.

PR 18-SEP-1992; 92US-00947338.

PR 17-SEP-1993; 93WO-US008610.

PR 17-APR-1995; 95US-00397232.

PR 07-JUN-1995; 95US-00475886.

PR 31-AUG-2000; 2000US-00653499.

XX (FUNK/) FUNKHOUSER A W.

PA (EMER/) EMERSON S U.

PA (PURC/) PURCELL R H.

PA (DHON/) D'HONDT E.

XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;

XX WPI; 2003-352605/02.

DR N-PSDB; ABX93474.

XX New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells, useful in vaccines for protecting primates against hepatitis infection and disease.

XX Example 3; Fig 6; 70pp; English.

XX The invention describes a live hepatitis A virus (HAV) adapted to growth in MRC-5 cells. The HAV and compositions comprising the HAV are useful as a vaccine for protecting primates against hepatitis infection and disease. This is the amino acid sequence of an attenuated (pass 35) human hepatitis A virus strain HM-175. (Updated on 23-OCT-2003 to standardise OS field)

XX SQ Sequence 2227 AA;

Query Match 100.0%; Score 101; DB 6; Length 2227;

Best Local Similarity 100.0%; Pred. No. 6.1e-07;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGLFSQAKISLF 20

Db 823 VLPPrKMKGLFSQAKISLF 842

RESULT 19

ABU08641

ID ABU08641 standard; protein; 2227 AA.

XX AC ABU08641;

XX 23-OCT-2003 (revised)

DT 03-JUN-2003 (first entry)

XX Attenuated hepatitis A virus (4380) strain HM-175.

XX Hepatitis A virus; HAV; virucide; hepatotropic; antiinflammatory; vaccine; MRC-5 cell; hepatitis infection; ds; attenuated virus.

XX Hepatitis A virus; strain HM-175.

XX US2002176869-A1.

PN 28-NOV-2002.

XX 29-APR-2002; 2002US-00135988.

PR 18-SEP-1992; 92US-00947338.

PR 17-SEP-1993; 93WO-US008610.

PR 17-APR-1995; 95US-00397232.

PR 07-JUN-1995; 95US-00475886.

PR 31-AUG-2000; 2000US-00653499.  
 XX (FUNK/) FUNKHOUSER A W.  
 PA (EMER/) EMERSON S U.  
 PA (PURC/) PURCELL R H.  
 PA (DHON/) D'HONDT E.  
 XX  
 PI Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
 XX  
 XX WPI; 2003-352605/02.  
 DR N-PSDB; ABX93475.  
 XX  
 XX New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells,  
 PT useful in vaccines for protecting primates against hepatitis infection  
 PT and disease.  
 XX  
 XX Disclosure; Page 45-51; 70pp; English.  
 XX  
 XX The invention describes a live hepatitis A virus (HAV) adapted to growth  
 CC in MRC-5 cells. The HAV and compositions comprising the HAV are useful as  
 CC a vaccine for protecting primates against hepatitis infection and  
 CC disease. This is the amino acid sequence of an attenuated human  
 CC hepatitis A virus (4380) strain HM-175. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 XX  
 XX Sequence 2227 AA;  
 SQ

Query Match 100.0%; Score 101; DB 6; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
 |||||  
 DB 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 20  
 ABU08639  
 ID ABU08639 standard; protein; 2227 AA.  
 XX  
 AC ABU08639;  
 XX  
 DT 23-OCT-2003 (revised)  
 DT 03-JUN-2003 (first entry)  
 XX  
 DE Wild type human hepatitis A virus strain HM-175.  
 XX  
 KW Hepatitis A virus; HAV; virucide; hepatotropic; antiinflammatory;  
 KW vaccine; MRC-5 cell; hepatitis infection.  
 XX  
 OS Hepatitis A virus; strain HM-175.  
 XX  
 PN US2002176869-A1.  
 XX  
 PD 28-NOV-2002.  
 XX  
 XX 29-APR-2002; 2002US-00135988.  
 XX  
 XX 18-SEP-1992; 92US-00947338.  
 PR 17-SEP-1993; 93WO-US008610.  
 PR 17-APR-1995; 95US-00397232.  
 PR 07-JUN-1995; 95US-00475886.  
 PR 31-AUG-2000; 2000US-00653499.  
 XX  
 PA (FUNK/) FUNKHOUSER A W.  
 PA (EMER/) EMERSON S U.  
 PA (PURC/) PURCELL R H.  
 PA (DHON/) D'HONDT E.  
 XX  
 XX Funkhouser AW, Emerson SU, Purcell RH, D'hondt E;  
 XX  
 XX WPI; 2003-352605/02.  
 DR N-PSDB; ABX93473.

XX New live hepatitis A virus, HAV4380, adapted to growth in MRC-5 cells,  
 PT useful in vaccines for protecting primates against hepatitis infection  
 PT and disease.  
 XX  
 XX Disclosure; Fig 6; 70pp; English.  
 XX  
 XX The invention describes a live hepatitis A virus (HAV) adapted to growth  
 CC in MRC-5 cells. The HAV and compositions comprising the HAV are useful as  
 CC a vaccine for protecting primates against hepatitis infection and  
 CC disease. This is the amino acid sequence of wild type human hepatitis A  
 CC virus strain HM-175. (Updated on 23-OCT-2003 to standardise OS field)  
 XX  
 XX Sequence 2227 AA;  
 SQ

Query Match 100.0%; Score 101; DB 6; Length 2227;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
 |||||  
 DB 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 21  
 ABW00350  
 ID ABW00350 standard; protein; 2227 AA.  
 XX  
 AC ABW00350;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Hepatitis A virus protein.  
 XX  
 KW Ribavirin; vaccine; immune response; infection; therapy; immunostimulant;  
 KW virucide.  
 XX  
 OS Hepatitis A virus.  
 XX  
 PN US2002136740-A1.  
 XX  
 PD 26-SEP-2002.  
 XX  
 XX 15-AUG-2001; 2001US-00929955.  
 PF 17-AUG-2000; 2000US-0225767P.  
 PR 29-AUG-2000; 2000US-0229175P.  
 XX  
 XX (SALL/) SALLBERG M.  
 PA (HULT/) HULTGREN C.  
 XX  
 PI Sallberg M, Hultgren C;  
 XX  
 XX WPI; 2003-764978/72.  
 DR N-PSDB; AAD60867.  
 XX  
 XX Vaccine compositions for treating and preventing disease, preferably  
 PT hepatitis C virus infection, comprises ribavirin and antigen that has  
 PT epitope present in hepatitis C virus.  
 XX  
 XX Claim 11; Page 45-51; Opp; English.  
 XX  
 XX The invention relates to a composition comprising ribavirin and an  
 CC antigen, where the antigen is derived from a hepatitis virus. The vaccine  
 CC is useful in enhancing the immune response to a hepatitis C antigen where  
 CC the composition is delivered to an animal identified as requiring an  
 CC enhanced immune response. The vaccine is useful in the treatment and  
 CC prevention of hepatitis C infection. The present sequence is Hepatitis A  
 CC virus protein  
 XX  
 XX Sequence 2227 AA;  
 SQ

Query Match 100.0%; Score 101; DB 7; Length 2227;

Best Local Similarity 100.0%; Pred. No. 6.1e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
 |||||  
 DB 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 22  
 AAP60066  
 ID AAP60066 standard; protein; 2227 AA.  
 XX AC AAP60066;  
 XX 25-MAR-2003 (revised)  
 XX 26-JUN-1991 (first entry)  
 XX DE Sequence of viral 1434 polypeptide encoded by the complete nucleotide  
 DE sequence of the HAV genome.  
 XX  
 XX Diagnosis; vaccine; passive immunotherapy.  
 KW  
 XX Hepatitis A virus.  
 OS

Key	Location/Qualifiers
FFH Region	1..245
FT FT Region	/label= P1.1A
FT FT Region	246..491
FT FT Region	/label= 1B
FT FT Region	492..836
FT FT Region	/label= 1C
FT FT Region	837..980
FT FT Region	/label= P2.2A
FT FT Region	981..1076
FT FT Region	/label= 2B
FT FT Region	1077..1422
FT FT Region	/label= 2C
FT FT Region	1423..1484
FT FT Region	/label= P3.3A
FT FT Region	1485..1507
FT FT Region	/label= 3B
FT FT Region	1508..1678
FT FT Region	/label= 3C
FT FT Region	1679..2227
FT FT Region	/label= 3D
XX	
PN	EP199480-A.
XX	
PD	29-OCT-1986.
XX	
PF	03-APR-1986; 86EP-00302465.
XX	
PR	03-APR-1985; 85US-00719329.
XX	
PA	(CHIR ) CHIRON CORP.
XX	
PI	Dina D, Potter SJ, Vannest GA, Caput D;
XX	
DR	WPI; 1986-286213/44.
XX	
DR	N-PSDB; AAN60080.
XX	
PT	Hepatitis A virus nucleotide sequence and polypeptide - and use in prodn.
XX	
PS	of vaccines and diagnostic probes.
XX	
XX	Claim 5; Fig 1; 18pp; English.
XX	
CC	AAN60080 and oligonucleotide fragments are useful in detection of
CC	hepatitis A virus; transformed hosts may be used for expression of
CC	polypeptides and fragments useful in vaccines without risk of infection
CC	by the virus or in prodn. of particles which are capable of inducing
CC	immunocompetent B cells for passive immunotherapy. Pref. epitope is
CC	derived from AAs 445-657 or 782-848 of the HAV polypeptide sequence
CC	(AAP60066). (Updated on 25-MAR-2003 to correct PA field.)

CC cyno-HAV. The other two cleavage sites are the same. Two residues have  
CC been identified as part of the immuno- dominant region (see feature  
CC table) and are different to those in the same position in human HAV. The  
CC protein and peptides derived from it can be used in the prepn. of  
CC vaccines for the prevention of HAV infection. See also AAR15056. (Note:  
CC Revised entry submitted to correct the patent number format of US  
CC Government-owned NTIS applications to prevent clashes with ongoing US  
CC granted patent numbers. For further information please visit the Derwent  
CC web site at [www.derwent.com/dwpi/updates/ntis-us.html](http://www.derwent.com/dwpi/updates/ntis-us.html).) (Updated on 25-  
CC MAR-2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS  
CC field.) (Updated on 24-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 839 AA;

Query Match 83.2%; Score 84; DB 2; Length 839;  
Best Local Similarity 83.3%; Pred. No. 0.00012;  
Matches 15; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKIS 18  
:|||||:|||||:|||||  
DB 822 ILPPPRKMKGLFSQSKIS 839

RESULT 24  
AAW42930  
ID AAW42930 standard; peptide; 25 AA.

XX AC AAW42930;  
XX DT 28-APR-1998 (first entry)  
XX DE Immunogenic Hepatitis A virus peptide YK-1665.  
XX KW Immunogenic peptide; immunogenic epitope; P2A protein; immune response;  
XX OS Synthetic.  
XX OS Hepatitis A virus.  
XX PN WO9740147-A1.  
XX PD 30-OCT-1997.  
XX PF 18-APR-1997; 97WO-US006891.  
XX PR 19-APR-1996; 96US-0015644P.  
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX PI Fields HA, Khudyakov YE;  
XX PI WPI; 1997-535831/49.  
XX DR WPI; 1997-535831/49.  
XX CC Immunogenic Hepatitis A Virus (HAV) peptide(s) - used to induce an immune  
XX response to HAV in a mammal or to detect the presence of antibodies  
XX against HAV in a mammal.  
XX PS Claim 18; Page 112; 140pp; English.

Peptides AAW42922-30 are immunogenic peptides corresponding to  
CC immunogenic epitopes of the Hepatitis A virus (HAV). The peptides are  
CC substantially similar to a portion of the amino acid sequence of the P2A  
CC protein of HAV corresponding to amino acids 792-980. Compositions  
CC containing the peptides can be used to induce an immune response to HAV  
CC in a mammal. The peptides can also be used to detect the presence of  
CC antibodies against HAV in mammalian serum. The peptides can also be used  
CC to make an antibody against HAV by administering the peptide to a mammal  
XX  
SQ Sequence 25 AA;

Query Match 64.4%; Score 65; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.0036;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLF 12  
:|||||:|||||  
DB 14 VLPPPRKMKGLF 25

RESULT 25  
AAB69447  
ID AAB69447 standard; peptide; 26 AA.  
XX AC AAB69447;  
XX DT 20-APR-2001 (first entry)  
XX DE Synthetic HAV P2A peptide, SEQ ID NO: 47.  
XX KW Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
XX OS antigen; major structural capsid polypeptide; HAV antibody detection.  
XX OS Hepatitis A virus.  
XX OS Synthetic.  
XX PN WO200105824-A2.  
XX PD 25-JAN-2001.  
XX PF 14-JUL-2000; 2000WO-US019267.  
XX PR 15-JUL-1999; 99US-0144412P.  
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX PI Fields HA, Khudyakov YE;  
XX PI WPI; 2001-112681/12.  
XX PT Synthetic peptides used as antigen sources for enzyme immunoassays  
XX PT detecting anti-hepatitis A virus and as vaccines.  
XX PS Claim 13; Page 98; 130pp; English.

The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of an antibody, to detect acute phase infection by detecting  
CC the peptide to an antibody, to detect acute phase convalescence in a  
CC IGM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy end  
CC of the peptides enhances the IGM antibody reactivity

XX Sequence 26 AA;

Query Match 64.4%; Score 65; DB 4; Length 26;  
Best Local Similarity 100.0%; Pred. No. 0.0038;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLF 12  
:|||||:|||||  
DB 14 VLPPPRKMKGLF 25

RESULT 26  
AAR86001  
ID AAR86001 standard; protein; 1025 AA.  
XX

AC AAR36001;  
 XX 04-JUN-1996 (first entry)  
 DT Human dihydropyrimidine dehydrogenase.  
 DE Bovine; liver; human; dihydropyrimidine dehydrogenase; DPD;  
 XX pyrimidine catabolism; 5,6-dihydropyrimidine; pyrimidine analogue;  
 KW fluoropyrimidine; anticancer drug; 5-fluorouracil; FURA; cancer;  
 KW frameshift mutation.  
 XX Homo sapiens.  
 OS W09528489-A1.  
 XX W09528489-A1.  
 XX 26-OCT-1995.  
 PD 13-APR-1995; 95WO-US004567.  
 XX 13-APR-1994; 94US-00227357.  
 PR (UABR-) UAB RES FOUND.  
 XX Diasio RB, Lu Z, Zhang R, Johnson M, Cheng X;  
 PI WPI; 1995-373803/48.  
 DR N-PSDB; AAT03133.  
 XX Novel dihydro:pyrimidine dehydrogenase gene - used to optimise 5-  
 PT fluoro:uracil doses given to cancer patients.  
 XX Claim 9; Page 140-56; 207pp; English.  
 PS This sequence represents human liver dihydropyrimidine dehydrogenase  
 CC (DPD). DPD catalyses the initial and rate limiting step in pyrimidine  
 CC catabolism, the reduction of pyrimidines to 5,6-dihydropyrimidines. DPD  
 CC is a complex enzyme consisting of two identical subunits, containing FMN,  
 CC FAD and iron-sulphur centers, and utilising NADPH as a cofactor. DPD has  
 CC also been shown to catalyse the reduction of various pyrimidine analogues  
 CC including the fluoropyrimidine anticancer drug 5-fluorouracil (FURA). Up  
 CC to 85% of administered FURA may be catabolised by DPD, and it therefore  
 CC governs the effectiveness of FURA as an anticancer drug. DPD genes or  
 CC fragments of them may be used in the detection of DPD in a sample, esp.  
 CC isolated from a cancer patient. According to the amount of DPD detected,  
 CC a therapeutically effective amount of FURA may be determined and  
 CC administered. DPD deficiency, leading to life-threatening toxicity on  
 CC exposure to FURA, in a human caused by a frameshift mutation may be  
 CC determined by means of a molecular biological assay to detect the  
 CC deletion of an A residue at codon 318 within the DPD-coding region  
 XX SQ Sequence 1025 AA;  
 Query Match 52.0%; Score 52.5; DB 2; Length 1025;  
 Best Local Similarity 63.2%; Pred. No. 15;  
 Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;  
 QY 2 LPPPRKMKGLFSQAKISLF 20  
 DB 176 LPPPRKMSSEAYS-AKIALF 193  
 RESULT 27  
 AAR91420  
 XX AAR91420 standard; protein; 1025 AA.  
 XX AAR91420;  
 XX 11-JUN-1996 (first entry)  
 DT Human dihydropyrimidine dehydrogenase.  
 DE Dihydropyrimidine dehydrogenase; DPD; DPD gene; probe; deficiency;  
 KW 5-fluorouracil; cytostatic; cancer; antibody.

XX Homo sapiens.  
 OS Key  
 XX Binding-site /label= NADP(H)\_binding\_site  
 FT Binding-site 335..351  
 FT Binding-site 472..482  
 FT Binding-site /label= FAD\_binding\_site  
 FT Binding-site 661..678  
 FT Binding-site /label= Uracil\_binding\_site  
 FT Binding-site 953..964  
 FT Binding-site /label= 4Fe-4S\_binding\_site  
 FT Binding-site 986..997  
 FT /label= 4Fe-4S\_binding\_site  
 XX W09608568-A2.  
 XX 21-MAR-1996.  
 XX 07-SEP-1995; 95WO-US012016.  
 XX 12-SEP-1994; 94US-00304309.  
 XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX Gonzalez FJ, Fernandez-Salguero P;  
 PI WPI; 1996-179943/18.  
 DR N-PSDB; AAT14077.  
 XX Pig and human dihydro:pyrimidine dehydrogenase (DPD) genes and probes -  
 PT useful for detection of DPD deficiencies and identification of humans at  
 PT risk of toxic reaction to 5-fluoro:uracil anti-cancer treatment.  
 XX Claim 7; Page 48-51; 78pp; English.  
 CC Human dihydropyrimidine dehydrogenase (DPD) (AAR91420) can be expressed as  
 CC a recombinant protein in a prokaryotic host cell by insertion of an  
 CC encoding cDNA sequence (see AAT14077) into a vector, transfection of host  
 CC cells, and culturing the cells in medium contg. uracil, FAD and FMN. DPD  
 CC is used to raise antibodies useful in competitive binding immunoassays  
 XX SQ Sequence 1025 AA;  
 Query Match 52.0%; Score 52.5; DB 2; Length 1025;  
 Best Local Similarity 63.2%; Pred. No. 15;  
 Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;  
 QY 2 LPPPRKMKGLFSQAKISLF 20  
 DB 176 LPPPRKMSSEAYS-AKIALF 193  
 RESULT 28  
 AAR93361  
 ID AAR93361 standard; protein; 1025 AA.  
 XX AAR93361;  
 XX 28-MAY-1999 (first entry)  
 DT Human DPD protein.  
 DE DPD; dihydropyrimidine-dehydrogenase; monoclonal antibody; Mab;  
 KW immunoassay reagent; cancer patient; treatment; antitumour agent;  
 KW 5-fluorouracil; affinity purification; toxicity.  
 XX Homo sapiens.  
 XX DE19837391-A1.  
 XX 25-FEB-1999.



PF 18-AUG-1998; 98DE-01037391.  
PR 22-AUG-1997; 97EP-00114630.  
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.  
XX Yoshihiko T, Hasegawa M;  
PI WPI; 1999-155202/14.  
XX N-PSDB; AAX22903.  
DR Monoclonal antibody specific for dihydropyrimidine dehydrogenase - for  
PT assessing patient response to 5-fluorouracil antitumor agents.  
XX  
PS Disclosure; Page 17-22; 34pp; German.  
XX  
CC This invention describes a monoclonal antibody (MAB) specific for  
CC dehydrogenase (DPD). This MAB is used as immunoassay  
CC reagents to identify a lack of DPD in a patient and to assess the  
CC sensitivity of cancer patients to treatment with antitumor agents of the  
CC 5-fluorouracil (5-FU) type. The MAB can also be used for affinity  
CC purification of DPD. DPD is involved in reduction of 5-FU (and related  
CC catabolites and derivatives) and lack of it is associated with increased  
CC toxicity of this type of antitumor agent. It has specific binding  
CC interaction. The MAB provide a sensitive and reliable test for DPD, which  
CC is simple, rapid and suitable for routine screening  
XX  
SQ Sequence 1025 AA;

Query Match 52.0%; Score 52.5; DB 2; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 15;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;  
QY 2 LPPPRKMKGLFSQAKISLF 20  
DB 176 LPPPRKMKSEAYS-AKIALF 193

RESULT 29  
ADB70177  
ID ADB70177 standard; protein; 524 AA.  
XX  
AC ADB70177;  
DT 04-DEC-2003 (first entry)  
XX  
DE C. neoformans amino acid sequence SEQ ID NO:3221.  
XX  
KW fungicide; gene therapy; infection.  
XX  
OS Cryptococcus neoformans.  
XX  
PN WO2003052076-A2.  
XX  
PD 26-JUN-2003.  
XX  
PF 17-DEC-2002; 2002WO-US040225.  
XX  
PR 17-DEC-2001; 2001US-0341261P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Zamudio C, Eroshkin AM;  
XX  
DR WPI; 2003-533017/50.  
DR N-PSDB; ADB69094.  
XX  
PT New nucleic acid, useful for preparing a composition for treating an  
PT infection caused by Cryptococcus neoformans.  
XX  
PS Claim 9; SEQ ID NO 3221; 136pp; English.  
XX  
CC The invention relates to a novel purified or isolated Cryptococcus

CC neoformans nucleic acid molecule comprising a sequence encoding a  
CC polypeptide comprising a sequence not given in the specification. A  
CC polynucleotide of the invention has fungicide activity, and may have a  
CC use in gene therapy. The nucleic acid is useful for preparing a  
CC composition for treating an infection caused by Cryptococcus neoformans.  
CC The present sequence represents a C. neoformans sequence of the  
CC invention. Note: The sequence data for this patent is not represented in  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
SQ Sequence 524 AA;

Query Match 47.5%; Score 48; DB 7; Length 524;  
Best Local Similarity 43.8%; Pred. No. 39;  
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
QY 2 LPPPRKMKGLFSQAKI 17  
DB 5 LPPPRKMKSEAYSQSQM 20

RESULT 30  
AAG06085  
ID AAG06085 standard; protein; 407 AA.  
XX  
AC AAG06085;  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 2730.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137538P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142052P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144684P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.

PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 08-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 13-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.

PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 23-OCT-1999; 99US-0162142P.

Query Match 46.5%; Score 47; DB 3; Length 407;  
Best Local Similarity 56.2%; Pred. NO. 43;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQAKIS 18  
|||:|:|:  
Db 99 PPRDMQDLFKQNVLS 114

Search completed: May 11, 2004, 13:37:32  
Job time : 51 secs

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:30:01 ; Search time 10.5 Seconds  
(without alignments)  
183.222 Million cell updates/sec

Title: US-09-171-432A-42

Perfect score: 101

Sequence: 1 VLPPPRKMKGLFSQAKISLF 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR 78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	101	100.0	852	1 GNNYHA	genome polyprotein
2	101	100.0	1358	2 A03905	genome polyprotein
3	101	100.0	2227	1 GNNYHM	genome polyprotein
4	101	100.0	2227	1 GNNYHR	genome polyprotein
5	101	100.0	2227	1 GNNYMK	genome polyprotein
6	96	95.0	2230	1 GNNYSA	genome polyprotein
7	95	94.1	2227	1 GNNYHB	genome polyprotein
8	84	83.2	839	1 GNNYS2	genome polyprotein
9	75	74.3	341	2 S04137	dihydropyrimidine
10	52.5	52.0	1025	2 A54718	unknown protein T2
11	47	46.5	414	2 D96838	bacteriophage gene
12	45	44.6	79	2 A01062	probable 60S ribos
13	45	44.6	321	2 B86423	ribonucleoside-dip
14	44	43.6	340	1 WMBES7	hypothetical prote
15	44	43.6	343	2 S74937	C4-dicarboxylase-b
16	44	43.6	369	2 S77299	hypothetical prote
17	44	43.6	775	2 T48957	transposase all756
18	43	42.6	189	2 A12534	methanol dehydroge
19	43	42.6	309	2 D75008	hypothetical prote
20	43	42.6	379	2 T49919	heat shock protein
21	43	42.6	389	2 S41748	hypothetical prote
22	42.5	42.1	105	2 B72598	probable membrane
23	42	41.6	109	2 S69307	hypothetical prote
24	42	41.6	165	2 S02053	ribosomal protein
25	42	41.6	178	2 T02335	hypothetical prote
26	42	41.6	834	2 F82673	hypothetical prote
27	41.5	41.1	304	2 H82984	hypothetical prote
28	41	40.6	83	2 T17809	conserved hypothet
29	41	40.6	159	2 C72210	

30	41	40.6	213	2 H86638	hypothetical prote
31	41	40.6	229	2 T19153	hypothetical prote
32	41	40.6	237	2 C84065	transcription regu
33	41	40.6	370	2 G37001	endoglucanase fami
34	41	40.6	561	2 G36752	unknown protein F2
35	41	40.6	622	2 T37257	hypothetical prote
36	41	40.6	897	2 T06540	polyribonucleotide
37	41	40.6	1040	2 A34695	axonal glycoprotei
38	41	40.6	1234	2 T30254	jumonji protein -
39	40.5	40.1	924	2 T25007	hypothetical prote
40	40.5	40.1	932	2 T25008	hypothetical prote
41	40	39.6	177	2 C64133	mercuric resistanc
42	40	39.6	238	2 C83240	probable transcrip
43	40	39.6	259	2 F64532	conserved hypothet
44	40	39.6	272	2 T49070	probable heat choc
45	40	39.6	292	2 S23239	hypothetical prote
46	40	39.6	300	2 A18885	hypothetical prote
47	40	39.6	310	2 A55053	endothelial monocy
48	40	39.6	310	2 B55053	endothelial monocy
49	40	39.6	371	2 T46089	thyroid transcript
50	40	39.6	372	2 S53724	thyroid-specific
51	40	39.6	372	2 S12002	thyroid nuclear fa
52	40	39.6	380	2 T20269	hypothetical prote
53	40	39.6	401	2 G02321	thyroid transcript
54	40	39.6	448	2 D64567	glutamate dehydrog
55	40	39.6	523	2 S53945	hypothetical prote
56	40	39.6	562	2 H69545	probable fatty-acid
57	40	39.6	619	2 T03143	R-transactivator p
58	40	39.6	620	2 T50232	actin-like protein
59	40	39.6	645	2 A12963	cellulose synthase
60	40	39.6	645	2 D93319	hypothetical prote
61	40	39.6	649	2 T01106	hypothetical prote
62	40	39.6	739	2 T12964	subtilisin homolog
63	40	39.6	887	2 B96598	hypothetical prote
64	40	39.6	1048	2 T31425	C-terminal domain-
65	39.5	39.1	1040	1 A38306	alpha-mannosidase
66	39	38.6	204	2 D71339	probable ribosomal
67	39	38.6	222	2 G97704	hypothetical prote
68	39	38.6	224	2 A95571	conserved hypothet
69	39	38.6	238	2 E83502	probable transcrip
70	39	38.6	252	2 AB1030	probable membrane
71	39	38.6	274	2 B84847	hypothetical prote
72	39	38.6	280	2 T18443	hypothetical prote
73	39	38.6	347	2 T07108	glucan endo-1,3-be
74	39	38.6	395	2 T32309	hypothetical prote
75	39	38.6	427	2 E72488	probable tryptopha
76	39	38.6	505	2 A5916	probable sugar ABC
77	39	38.6	567	2 C85643	hypothetical prote
78	39	38.6	567	2 H90782	hypothetical prote
79	39	38.6	631	2 A57286	probable serine/th
80	39	38.6	651	2 T31513	hypothetical prote
81	39	38.6	675	2 PQ0227	adenylate cyclase
82	39	38.6	721	2 JC7557	lipidosis - mouse
83	39	38.6	736	2 T12963	subtilisin homolog
84	39	38.6	753	2 T01619	hypothetical prote
85	39	38.6	941	2 AF2415	two-component hybr
86	39	38.6	1127	2 E97580	hypothetical prote
87	39	38.6	1248	2 A53588	adenylate cyclase
88	39	38.6	1251	2 S48687	type VIII adenylyl
89	39	38.6	1358	2 T40737	probable calcium-t
90	39	38.6	1387	2 JCS502	G-protein signalin
91	38.5	38.1	585	2 S55205	dihydroxy-acid deh
92	38.5	38.1	881	2 S67026	probable membrane
93	38	37.6	49	2 F69412	hypothetical prote
94	38	37.6	156	2 B44048	hypothetical 18.1K
95	38	37.6	186	2 T50403	probable succinate
96	38	37.6	190	2 G70231	conserved hypothet
97	38	37.6	192	2 S76867	hypothetical prote
98	38	37.6	233	2 E97120	ribosomal protein
99	38	37.6	248	4 I59064	transforming prote
100	38	37.6	273	2 D69633	glutamine ABC tran

## ALIGNMENTS

C&gt;Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999

C:Accession: A25981  
R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.  
J. Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with di  
A:Reference number: A25981; MUID:87061253; PMID:3023706  
A:Accession: A25981  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M14707; NID:G329582; PIDN:AAA45465.1; PID:G329583  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-491/Product: coat protein 1C #status predicted <VP3>  
F:492-980/Product: coat protein 1D #status predicted <VP1>  
F:981-1087/Product: coat protein 2A #status predicted <C2A>  
F:1088-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1496/Product: core protein 3A #status predicted <C3A>  
F:1497-1519/Product: core protein 3B #status predicted <C3B>  
F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;

Best Local Similarity 100.0%; Pred. No. 3.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||  
DB 823 VLPPPRKMKGLFSQAKISLF 842  
|||

## RESULT 4

GNNYHR  
genome polyprotein - human hepatitis A virus  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core pr  
NA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
R:Accession: A03903  
R:Najarian, R.; Caput, D.; Gee, W.; Potter, S.J.; Renard, A.; Merryweather, J.; van Nest  
Proc. Natl. Acad. Sci. U.S.A. 82, 2627-2631, 1985  
A:Title: Primary structure and gene organization of human hepatitis A virus.  
A:Reference number: A03903; MUID:85190549; PMID:2866127  
A:Accession: A03903  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <NAJ>  
A:Cross-references: GB:K02990; NID:G329596; PIDN:AAA45472.1; PID:G329597  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-980/Product: core protein 2A #status predicted <C2A>  
F:981-1076/Product: core protein 2B #status predicted <C2B>  
F:1077-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1484/Product: protein 3A #status predicted <C3A>  
F:1485-1507/Product: protein 3B #status predicted <C3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;

Best Local Similarity 100.0%; Pred. No. 3.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||  
DB 823 VLPPPRKMKGLFSQAKISLF 842  
|||

## RESULT 5

## ALIGNMENTS

C&gt;Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999

C:Accession: A25981  
R:Cohen, J.I.; Ticehurst, J.R.; Purcell, R.H.; Buckler-White, A.; Baroudy, B.M.  
J. Virol. 61, 50-59, 1987  
A:Title: Complete nucleotide sequence of wild-type hepatitis A virus: comparison with di  
A:Reference number: A25981; MUID:87061253; PMID:3023706  
A:Accession: A25981  
A:Molecule type: genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M14707; NID:G329582; PIDN:AAA45465.1; PID:G329583  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans  
F:1-23/Product: coat protein 1A #status predicted <VP4>  
F:24-245/Product: coat protein 1B #status predicted <VP2>  
F:246-491/Product: coat protein 1C #status predicted <VP3>  
F:492-980/Product: coat protein 1D #status predicted <VP1>  
F:981-1087/Product: coat protein 2A #status predicted <C2A>  
F:1088-1422/Product: core protein 2C #status predicted <C2C>  
F:1423-1496/Product: core protein 3A #status predicted <C3A>  
F:1497-1519/Product: core protein 3B #status predicted <C3B>  
F:1520-1738/Product: cysteine proteinase, protein 3C #status predicted <C3C>  
F:1739-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <C3D>

Query Match 100.0%; Score 101; DB 1; Length 852;

Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||  
DB 823 VLPPPRKMKGLFSQAKISLF 842  
|||

## RESULT 2

A03905  
genome polyprotein (version 2) - human hepatitis A virus (fragments)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; protein  
C:Species: human hepatitis A virus  
C:Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 15-Nov-1996  
C:Accession: A03905  
R:Baroudy, B.M.; Ticehurst, J.R.; Miele, T.A.; Maizel Jr., J.V.; Purcell, R.H.; Feinsto  
Proc. Natl. Acad. Sci. U.S.A. 82, 2143-2147, 1985  
A:Title: Sequence analysis of hepatitis A virus cDNA coding for capsid proteins and RNA  
A:Reference number: A03905; MUID:8516289; PMID:2984684  
A:Accession: A03905  
A:Molecule type: genomic RNA  
A:Residues: 1-1358 <BAR>  
C:Superfamily: hepatitis A virus genome polyprotein  
C:Keywords: coat protein; core protein; polyprotein  
F:1-245/Product: coat protein 1A #status predicted <C1A>  
F:246-491/Product: coat protein 1B #status predicted <C1B>  
F:492-836/Product: coat protein 1C #status predicted <C1C>  
F:837-854/Product: core protein 2A (fragment) #status predicted <C2A>  
F:855-1358/Product: protein 3D (RNA polymerase) (fragment) #status predicted <C3D>

Query Match 100.0%; Score 101; DB 2; Length 1358;

Best Local Similarity 100.0%; Pred. No. 2.1e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||  
DB 823 VLPPPRKMKGLFSQAKISLF 842  
|||

## RESULT 3

GNNYHM  
genome polyprotein - human hepatitis A virus (strain HM-175, wild type)  
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro  
B, RNA-directed RNA polymerase (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A:Note: host Homo sapiens (man)

genome polypeptide - human hepatitis A virus (strain HM-175/7MK-5, attenuated HAV)  
NA contains: coat protein 1A; coat protein 1B; coat protein 1C; core protein 2A; core pro  
Nucleotide sequence (EC 2.7.7.48), protein 3D  
C:Species: human hepatitis A virus  
A>Note: host Homo sapiens (man)  
C>Date: 30-Jun-1988 #sequence revision 30-Jun-1988 #text\_change 16-Jul-1999  
C:Accession: A94149; A25914; A94508  
R:Cohen, J.I.; Rosenblum, B.; Tiechurst, J.R.; Daemer, R.J.; Feinstein, S.M.; Purcell, R.  
Proc. Natl. Acad. Sci. U.S.A. 84, 2497-2501, 1987  
A>Title: Complete nucleotide sequence of an attenuated hepatitis A virus: comparison wit  
A:Reference number: A94149; MUID:87175701; PMID:3031686  
A:Accession: A94149  
A>Status: nucleic acid sequence not shown  
A:Molecule type: Genomic RNA  
A:Residues: 1-2227 <COH>  
A:Cross-references: EMBL:M1632; NID:g329594; PID:AAA45471.1; PID:g329595  
A>Note: submitted to GenBank, August 1987  
C:Superfamily: hepatitis A virus genome polypeptide  
C:Keywords: coat protein; core protein; cysteine proteinase; hydrolase; nucleotidyltrans  
F:1-245/Product: coat protein 1A #status predicted <P1A>  
F:246-491/Product: coat protein 1B #status predicted <P1B>  
F:492-836/Product: coat protein 1C #status predicted <P1C>  
F:837-980/Product: core protein 2A #status predicted <P2A>  
F:981-1076/Product: core protein 2B #status predicted <P2B>  
F:1077-1422/Product: core protein 2C #status predicted <P2C>  
F:1423-1484/Product: protein 3A #status predicted <P3A>  
F:1485-1507/Product: protein 3B #status predicted <P3B>  
F:1508-1678/Product: cysteine proteinase, protein 3C #status predicted <P3C>  
F:1679-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 3.7e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

823 VLPPPRKMKGLFSQAKISLF 842

GENVISA

genome polyprotein - simian hepatitis A virus (strain AGM-27)  
A:NC\_009683; Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core protein  
A:CP; Species: simian hepatitis A virus  
A:C-C-Date: 31-Mar-1992 #sequence revision 31-Mar-1992 #text\_change 16-Jun-2000  
A:C-C-Accession: A30470; S04885; S03965  
R:Tsaarev, S.A.  
A:Reference number: J1P1D, April 1991  
A:Accession: A30470  
A:Reference number: A30470  
A:Molecule type: Genomic RNA  
A:Residues: 1-2230 <TSA>  
A:Cross-references: GB:D00924; NID:g222597; PIDN:BAAC00766.1; PID:g222598  
R:Tsaarev, S.A.; Emerson, S.U.; Balaayan, M.S.; Ticehurst, J.; Purcell, R.H.  
J. Gen. Virol. 72, 1677-1683, 1991  
A>Title: Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure an  
A:Reference number: JQ1080; UID:91311420; PMID:1645901  
A:Contents: annotation  
A:Nore: neither amino acid nor nucleotide sequence is given  
S:Balaayan, M.S.; Kusov, Y.Y.; Andjaparidze, A.G.; Tsaarev, S.A.; Chizhik  
submitted to the EMBL Data Library, May 1989  
A:Reference number: S04885  
A:Accession: S04885

A.Molecule type: GENOMIC RNA  
A.Residues: 1750-2164 <BALL>  
A.Cross-references: ENML:X15461; PID: g930268  
A.Author: R.R.Balayán, M.S.: Kusov, Y.Y.; Andjaparidze, A.G.; Tearev, S.A.; Sverdlov, E.D.; Chizhikov, A.G.; Title: Variations in genome fragments coding for RNA polymerase in human and simian he  
A.Reference number: S03965; PMID: 89232168; PMID: 2541023  
A.Accession: S03965  
A.Molecule type: GENOMIC RNA

Virus Res. 8, 153-171, 1987  
 Krieger, A.V., Tada, H., von der Helm, K., Wissel, T., Klenh, K., Wamter, B., Deinhardt, K.  
 A:Title: The entire nucleotide sequence of the genome of human hepatitis A virus (isolat  
 A:Reference number: JS0303; MUID:88045071; PMID:2823500  
 A:Accession: JS0303

A.Cross-references: EMBL:M20273

C.Superfamily: hepatitis A virus genome polyprotein

C.Keywords: coat protein; coat protein; cysteine proteinase; genome-linked protein; hydr

F.1-23/Product: coat protein 1A #status predicted <VP4>

F.24-246/Product: coat protein 1B #status predicted <VP2>

F.247-491/Product: coat protein 1C #status predicted <VP3>

F.492-836/Product: coat protein 1D #status predicted <VP1>

F.837-980/Product: coat protein 2A #status predicted <P2A>

F.981-1108/Product: coat protein 2B #status predicted <P2B>

F.1109-1438/Product: coat protein 2C #status predicted <P2C>

F.1439-1496/Product: protein 3A #status predicted <P3A>

F.1497-1519/Product: genome-linked protein vpg #status predicted <VPG>

F.1520-1736/Product: cysteine proteinase, protein 3C #status predicted <P3C>

F.1737-2227/Product: RNA-directed RNA polymerase, protein 3D #status predicted <P3D>

Query Match 94.1%; Score 95; DB 1; Length 2227;

Best Local Similarity 95.0%; Pred. No. 3.7e-07;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

22 1 VLPPPRKMGKLFQAKISLF 20

db 823 VLPPPRKKKGLFQAKISLF 842

**RESULT 8**

CNNYS2  
genome polypeptide - simian hepatitis A virus (strain CY-145) (fragment)  
Contains: amino end of core protein 2A; coat protein 1A; coat protein 1B; coat protein C; Species: simian hepatitis A virus  
Host Macaca fascicularis (cynomolgus macaque)  
Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 16-Jul-1999  
Accession: J01180

R;Nainan, O.V.; Margolis, H.S.; Robertson, B.H.; Balayan, M.; Brinton, M.A.  
J. Gen. Virol. 72, 1685-1689, 1991  
A;Title: Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus mac  
A;Reference number: JQ1180; MUID:91311421; PMID:1649902  
A;Accession: JQ1180  
A;Molecule type: genomic RNA  
A;Residues: 1-839 <NAI>  
A;Cross-references: GB:M59286; NID:9329599; PIDN:AAA45473.1; PID:G555083  
C;Superfamily: Hepatitis A virus genome polyprotein  
C;Keywords: coat protein; core protein; glycoprotein; polypeptide  
F;1-23/Product: coat protein 1A #status predicted <VP0>  
F;24-245/Product: coat protein 1B #status predicted <VP3>  
F;246-491/Product: coat protein 1C #status predicted <VP1>  
F;492-839/Product: core protein 2A (fragment) #status predicted <P2P>  
F;261.312.728.756/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 83.2%; Score 84; DB 1; Length 839;  
Best Local Similarity 83.3%; Pred. No. 8.7e-06;  
Matches 15; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKIS 18  
:|||||:|||||:  
Db 822 ILPPPRKMKGLFSQAKIS 839

RESULT 9  
S04137  
genome polyprotein - human hepatitis A virus (strain LCDC-1) (fragment)  
C;Species: human hepatitis A virus  
C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
C;Accession: S04137  
R;Andonov, A.P.; Lau, P.C.K.; Chaudhary, R.  
Nucleic Acids Res. 17, 3594, 1989  
A;Title: Nucleotide sequence of the VP1 gene from a Chinese strain of hepatitis A virus  
A;Reference number: S04137; MUID:89263805; PMID:2542903  
A;Accession: S04137  
A;Molecule type: mRNA  
A;Residues: 1-341 <AND>  
A;Cross-references: EMBL:X14666; NID:G62301; PIDN:CAA32794.1; PID:G4377576  
C;Gene: VPI  
C;Superfamily: hepatitis A virus genome polyprotein  
C;Keywords: coat protein; polypeptide  
F;2-340/Product: coat protein 1D (VPI) #status predicted <WAT>

Query Match 74.3%; Score 75; DB 2; Length 341;  
Best Local Similarity 93.3%; Pred. No. 0.0001;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQA 15  
:|||||:|||||:  
Db 327 VLPPPRKMKGLFSQS 341

RESULT 10  
A54718  
dihydropyrimidine dehydrogenase (NADP) (EC 1.3.1.2) - human  
C;Species: Homo sapiens (man)  
C;Date: 28-Apr-1995 #sequence\_revision 28-Apr-1995 #text\_change 03-Jun-2002  
C;Accession: A54718  
R;Yokota, H.; Fernandez-Salguero, P.; Furuya, H.; Lin, K.; McBride, O.W.; Podschun, B.;  
J. Biol. Chem. 269, 23192-23196, 1994  
A;Title: cDNA cloning and chromosome mapping of human dihydropyrimidine dehydrogenase, a  
A;Reference number: A54718; MUID:94365020; PMID:8083224  
A;Accession: A54718  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1025 <YOK>  
A;Cross-references: GB:U09178; NID:G558304; PID:G558305  
A;Note: authors translated the codon GGC for residue 748 as Asp, and TTA for residue 101  
C;Gene: GDB:DPYD  
A;Cross-references: GDB:364102; OMIM:274270

A;Map position: lp22-1p22  
C;Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S] ho  
C;Keywords: 4Fe-4S; flavoprotein; homodimer; iron-sulfur protein; NADP; oxidoreductase  
F;946-1005/Domain: ferredoxin 2[4Fe-4S] homology <FER>

Query Match 52.0%; Score 52.5; DB 2; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 2.1;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 2 LPPPRKMKGKGLFSQAKISLF 20  
|||||:|||||:  
Db 176 LPPPRKMSEAYS-AKIALF 193

RESULT 11  
D96838  
unknown protein T21F11.5 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: D96838  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, T.H.; Dewar, K.;  
ansen, N.P.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: AB6141; MUID:21016719; PMID:11130712  
A;Accession: D96838  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-414 <STO>  
A;Cross-references: GB:AB005173; NID:G6730725; PIDN:AAF27115.1; GSPDB:GN00141  
C;Gene: T21F11.5  
A;Map position: 1

Query Match 46.5%; Score 47; DB 2; Length 414;  
Best Local Similarity 56.2%; Pred. No. 6.2;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 PPFRKMKGLFSQAKIS 18  
|||||:|||||:  
Db 106 PPFRKMKGLFSQAKIS 121

RESULT 12  
AG1062  
bacteriophage gene regulatory protein STY4826 [imported] - Salmonella enterica subsp. en  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AG1062  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,  
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gara, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AG1062  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-79 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAB06948.1; PID:G16505595; GSPDB:GN00176  
C;Gene: STY4826

Query Match 44.6%; Score 45; DB 2; Length 79;  
Best Local Similarity 63.6%; Pred. No. 2.1;

```

Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 PPPRKMKGFLS 13
Db 64 PPPRKMKGFLS 74

RESULT 13
E86423
probable 60S ribosomal protein L18A - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C:Accession: E86423
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, B.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E86423
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-321 <STO>
A:Cross-references: GB:AE005172; NID:G10092460; PIDN:AG12862.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 44.6%; Score 45; DB 2; Length 321;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGFLSQAKISLF 20
Db 301 VLPPrKLTTPKANKPNLF 320

RESULT 14
WMB57
ribonucleoside-diphosphate reductase (EC 1.17.4.1) small chain - human herpesvirus 1 (st
N:Alternate names: ribonucleotide reductase small chain
C:Species: human herpesvirus 1
C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 15-Sep-2003
C:Accession: D30088
R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perz
J. Gen. Virol. 69, 1531-1574, 1988
A:Title: The complete DNA sequence of the long unique region in the genome of herpes sim
A:Reference number: A30083; MUID:88274327; PMID:2839594
A:Accession: D30088
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-340 <MCG>
A:Cross-references: GB:X14112; NID:G1944536; PIDN:CAA32303.1; PID:G59540; GB:D00317
C:Genetics:
A:Gene: UL40
C:Function:
A:Description: catalyzes the reduction of ribonucleoside diphosphate to deoxyribonucleos
C:Superfamily: ribonucleoside reductase small subunit
C:Keywords: deoxyribonucleotide biosynthesis; early protein; iron; metalloprotein; oxido
F;94,124,127,187,221,224/Binding site: 2Fe-O cluster (Asp, Glu, His, Glu, Glu, His) #sta
F;131/Active site: Tyr (stable tyrosyl radical) #status predicted

Query Match 43.6%; Score 44; DB 1; Length 340;
Best Local Similarity 61.5%; Pred. No. 16;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PPPRKMKGFLSQA 15
Db 242 PPPRKYGLFQA 254

```

## RESULT 15

S74937

hypothetical protein slr0688 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C:Accession: S74937

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

S

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S74937

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-343 &lt;KAN&gt;

A:Cross-references: EMBL:D90902; GB:AB001339; NID:G1652027; PIDN:BAA16977.1; PID:dl01771

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 43.6%; Score 44; DB 2; Length 343;

Best Local Similarity 47.1%; Pred. No. 16;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PPPRKMKGFLSQAKISL 19

Db 240 PDPRECKGLFRGAFTI 256

## RESULT 16

S77299

C4-dicarboxylase-binding protein - Synechocystis sp. (strain PCC 6803)

N:Alternate names: protein sl1314

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C:Accession: S77299

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

S

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S77299

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-369 &lt;KAN&gt;

A:Cross-references: EMBL:D90907; GB:AB001339; NID:G1652618; PIDN:BAA17633.1; PID:dl01836

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: dctp

Query Match 43.6%; Score 44; DB 2; Length 369;

Best Local Similarity 50.0%; Pred. No. 17;

Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 1 VLPPrKMKGFLSQAKISLF 20

Db 38 ILPLPGKGYYSQAKIRAF 57

## RESULT 17

T48957

hypothetical protein T15B3.170 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000

C:Accession: T48957

R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.; I

submitted to the Protein Sequence Database, April 2000

A:Reference number: 225009

A:Accession: T48957



A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-775 <JUR>  
A;Cross-references: EMBL:AL163975; GSPDB:GN00061; ATSP:T15B3.170  
A;Experimental source: cultivar Columbia; BAC clone T15B3  
C;Genetics:  
A;Gene: ATSP:T15B3.170  
A;Map position: 3  
A;Introns: 33/3; 135/3; 298/1; 376/2; 478/3; 551/3

Query Match 43.6%; Score 44; DB 2; Length 775;  
Best Local Similarity 52.9%; Pred. No. 40;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQAKISL 19  
Db 749 PPRKMKHLPNQCHTSL 765

RESULT 18  
AI2534  
transposase all7564 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120beta  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C;Accession: AI2534  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AI2534  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-189 <KUR>  
A;Cross-references: GB:AP003602; PIDN:BAE77207.1; PID:g17134649; GSPDB:GN00181  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: all7564  
A;Genome: plasmid

Query Match 42.6%; Score 43; DB 2; Length 189;  
Best Local Similarity 55.6%; Pred. No. 12;  
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 LPPPRKMKGLFSQAKISL 19  
Db 162 LTPPLKMKNLQWRKILL 179

RESULT 19  
D75008  
methanol dehydrogenase regulatory protein (moxr-2) PAB1296 - Pyrococcus abyssi (strain C  
C;Species: Pyrococcus abyssi  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C;Accession: D75008  
R;anonymus; Genoscope  
submitted to the EMBL Data Library, July 1999  
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str  
A;Reference number: A75001  
A;Accession: D75008  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-309 <KAW>  
A;Cross-references: GB:AJ248288; GB:AL096836; NID:95458960; PIDN:CAB50506.1; PID:g545902  
A;Experimental source: strain Orsay  
C;Genetics:  
A;Gene: PAB1296  
C;Superfamily: methanol dehydrogenase regulatory protein

Query Match 42.6%; Score 43; DB 2; Length 309;  
Best Local Similarity 81.8%; Pred. No. 21;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 6 RKMKGFLFSQAK 16  
Db 185 RKDKGLFSQAK 195

RESULT 20  
T49919  
hypothetical protein F17114.20 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C;Accession: T49919  
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z24490  
A;Accession: T49919  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-379 <BEV>  
A;Cross-references: EMBL:AL353994; GSPDB:GN00063; ATSP:F17114.20  
A;Experimental source: cultivar Columbia; BAC clone F17114  
C;Genetics:  
A;Gene: ATSP:F17114.20  
A;Map position: 5  
A;Introns: 119/1; 146/1; 222/3; 256/2; 327/2

Query Match 42.6%; Score 43; DB 2; Length 379;  
Best Local Similarity 77.8%; Pred. No. 26;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 PPRKMKGL 11  
Db 32 PPRKMKSM 40

RESULT 21  
S41748  
heat shock protein dnaJ - Methanosarcina mazei  
C;Species: Methanosarcina mazei  
C;Date: 31-Dec-1993 #sequence\_revision 02-Aug-1994 #text\_change 20-Aug-1999  
C;Accession: S41748; S41150  
R;Conway de Macario, E.  
submitted to the EMBL Data Library, April 1992  
A;Reference number: S41748  
A;Accession: S41748  
A;Molecule type: DNA  
A;Residues: 1-389 <CON>  
A;Cross-references: EMBL:X60265; NID:g48938; PIDN:CAA42813.1; PID:g48940  
A;Experimental source: strain S-6  
R;Macario, A.J.L.; Dugan, C.B.; Clarens, M.; Conway de Macario, E. Nucleic Acids Res. 21, 2773, 1993  
A;Title: dnaJ in Archaea.  
A;Reference number: S41150; MUID:93324351; PMID:8332479  
A;Accession: S41150  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 4-208 <MAC>  
A;Cross-references: EMBL:X60265; NID:g48938; PIDN:CAA42813.1; PID:g48940  
A;Experimental source: strain S-6  
C;Genetics:  
A;Gene: dnaJ  
C;Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
C;Keywords: DNA replication; heat shock; molecular chaperone; stress-induced protein  
F;6-70/Domain: dnaJ amino-terminal homology <DNJ>  
F;77-104/Region: G/P motif  
F;144-151/Region: CXXCXGKG repeat  
F;161-168/Region: CXXCXGKG repeat  
F;187-194/Region: CXXCXGKG repeat  
F;201-208/Region: CXXCXGKG repeat

Query Match 42.6%; Score 43; DB 2; Length 389;  
Best Local Similarity 66.7%; Pred. No. 27;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 5 PRKMKGLFSQAK 16
   ||| ||| :
Db 373 PRKSGLFPEVK 384

RESULT 22
B72598
Hypothetical protein APE1250 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: B72598
R:Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339; PMID:10382956
A:Accession: B72598
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-105 <KAW>
A:Cross-references: DBJ:AP000061; NID:G5104821; PIDN:BA080240.1; PID:d1044026; PID:G510
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1250

Query Match 42.1%; Score 42.5; DB 2; Length 105;
Best Local Similarity 61.1%; Pred. NO. 7.6;
Matches 11; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 VLPPPRKMKGLFSQAKIS 18
   ||| ||| :
Db 36 VLPPPRAM-SMASQGLIS 52

RESULT 23
S69307
Probable membrane protein YLR294c - yeast (Saccharomyces cerevisiae)
A:Alternate names: hypothetical protein I8003.19-a
C:Species: Saccharomyces cerevisiae
C:Date: 20-Jul-1996 #sequence_revision 23-Aug-1996 #text_change 19-Apr-2002
C:Accession: S69307
R:Pauley, A.
submitted to the EMBL Data Library, November 1994
A:Description: The sequence of S. cerevisiae cosmid 8003.
A:Reference number: S50366
A:Accession: S69307
A:Molecule type: DNA
A:Residues: 1-109 <PAU>
A:Cross-references: EMBL:U17243; NID:G596030; PID:G2340967; GSPDB:GN00012; MIPS:YLR294c
C:Genetics:
A:Gene: MIPS:YLR294c
A:Cross-references: SGD:S0004285
A:Map position: 12R
C:Superfamily: Saccharomyces probable membrane protein YLR294c
C:Keywords: transmembrane protein
P:77-93/Domain: transmembrane #status predicted <TM>

Query Match 41.6%; Score 42; DB 2; Length 109;
Best Local Similarity 50.0%; Pred. NO. 9.7;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 PRKMKGLFSQAKISLF 20
   ||| ||| :
Db 63 PHKQNLFFKQKIQLY 78

RESULT 24
S02053
Hypothetical protein (D10 5' region) - phage T5
C:Species: phage T5
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 08-Oct-1999
C:Accession: S02053
```

```
R:Kallman, A.V.; Kryukov, V.M.; Bayev, A.A.
Nucleic Acids Res. 16, 10353-10354, 1988
A:Title: The nucleotide sequence of the region of bacteriophage T5 early genes D10-D15.
A:Reference number: S01931; MUID:89057468; PMID:3057441
A:Accession: S02053
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-165 <KAL>
A:Cross-references: EMBL:X12930; NID:G15407; PIDN:CAA31397.1; PID:G579170
C:Genetics:
A:Start codon: GTG
C:Superfamily: phage T5 hypothetical protein (D10 5' region)

Query Match 41.6%; Score 42; DB 2; Length 165;
Best Local Similarity 53.3%; Pred. NO. 15;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 6 RKMKGGLFSQAKISLF 20
   ||| ||| :
Db 38 RKLPGTFSQRRLILF 52

RESULT 25
T02335
Ribosomal protein L18a, cytosolic - Arabidopsis thaliana
N:Alternate names: protein F13P17.31
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02335; B84757
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC F13P17 genomic sequence.
A:Reference number: Z14657
A:Accession: T02335
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-178 <ROU>
A:Cross-references: EMBL:AC004481; NID:G3337347; PIDN:AAC27421.1; PID:G3337376
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84757
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-178 <STO>
A:Cross-references: GB:AE002093; NID:G3128228; PIDN:AAC26708.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g34480; F13P17.31
A:Map position: 2
A:Introns: 5/3; 44/2; 65/3
C:Superfamily: rat ribosomal protein L18a
C:Keywords: cytosol; protein biosynthesis; ribosome

Query Match 41.6%; Score 42; DB 2; Length 178;
Best Local Similarity 45.0%; Pred. NO. 17;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20
   ||| ||| :
Db 158 VRPPSRKLTYYKANKPNLF 177

RESULT 26
F82673
Hypothetical protein XF1508 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82673
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
```

Nature 406, 151-157, 2000  
 A>Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A>Note: for a complete list of authors see reference number A59328 below  
 A:Accession: H82673  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-834 <SIM>  
 A:Cross-references: GB:AB003980; GB:AB003849; NID:G9106531; PIDN:AAF84317.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, W.R.F.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kasper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laigz  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF1508

Query Match 41.6%; Score 42; DB 2; Length 834;  
 Best Local Similarity 60.0%; Pred. No. 93;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 VLPFRKMKGLFSQA 15  
 DB 597 VAPPPRFVKTLRQA 611

RESULT 27  
 H82984  
 hypothetical protein PA5284 [imported] - *Pseudomonas aeruginosa* (strain PAO1)  
 C:Species: *Pseudomonas aeruginosa*  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: H82984  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lin,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337; PMID:10984043  
 A:Accession: H82984  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-304 <STO>  
 A:Cross-references: GB:AE004941; GB:AE004091; NID:G9951596; PIDN:AG08669.1; GSPDB:GN001  
 A:Experimental source: strain PAO1  
 C:Genetics:  
 A:Gene: PA5284

Query Match 41.1%; Score 41.5; DB 2; Length 304;  
 Best Local Similarity 50.0%; Pred. No. 37;  
 Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;  
 QY 2 LPPPRKMKGLFSQ-AKISLF 20  
 DB 102 IPASSRFKGTFSQIAEILF 121

RESULT 28  
 T17809  
 hypothetical protein a311R - *Chlorella virus* PBCV-1  
 C:Species: *Chlorella virus* PBCV-1  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T17809  
 R:Graves, M.V.; Van Etten, J.L.

submitted to the EMBL Data Library, May 1999  
 A:Reference number: Z18806  
 A:Accession: T17809  
 A>Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-83 <GRA>  
 A:Cross-references: EMBL:U42580; NID:G4028896; PIDN:AA066679.1  
 A:Experimental source: specific host *Chlorella* strain NC64A  
 C:Genetics:  
 A>Note: a311R

Query Match 40.6%; Score 41; DB 2; Length 83;  
 Best Local Similarity 47.1%; Pred. No. 10;  
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 LPPPRKMKGLFSQAKIS 18  
 DB 50 LPVPAIDKGFASAKLT 66

RESULT 29  
 C72210  
 conserved hypothetical protein - *Thermotoga maritima* (strain MSB8)  
 C:Species: *Thermotoga maritima*  
 C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C:Accession: C72210  
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hicke  
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
 C.M.  
 Nature 399, 323-329, 1999  
 A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
 A:Reference number: A72200; MUID:99287316; PMID:10360571  
 A:Accession: C72210  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <ARN>  
 A:Cross-references: GB:AE001817; GB:AE000512; NID:G4982370; PIDN:AA036861.1; PID:G498237  
 A:Experimental source: strain MSB8  
 C:Genetics:  
 A:Gene: TM1798  
 C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0377

Query Match 40.6%; Score 41; DB 2; Length 159;  
 Best Local Similarity 44.4%; Pred. No. 22;  
 Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 3 PPRPRKMKGLFSQAKISLF 20  
 DB 139 PPPPRKKGICRKGVELF 156

RESULT 30  
 H86638  
 hypothetical protein ybbe [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL1403)  
 C:Species: *Lactococcus lactis* subsp. *lactis*  
 C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: H86638  
 R:Botot, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrli  
 Genome Res. 11, 731-753, 2001  
 A>Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* s  
 A:Reference number: A86625; MUID:21235186; PMID:11337471  
 A:Accession: H86638  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-213 <STO>  
 A:Cross-references: GB:AE005176; PID:G12722959; PIDN:AA04210.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: ybbe

Query Match 40.6%; Score 41; DB 2; Length 213;  
 Best Local Similarity 46.7%; Pred. No. 30;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 PRKMGFLFSQAKISL 19  
| : : : :  
Db 153 PLTMKGVPKQKYSI 167

Search completed: May 11, 2004, 13:41:21  
Job time : 13.75 secs

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:24:26 ; Search time 6.75 Seconds  
(without alignments)  
154,282 Million cell updates/sec

Title: US-09-171-432A-42

Perfect score: 101

Sequence: 1 VLPPPRKMKGLFSQAKISLF 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	852	1	POLG HPACV
2	101	100.0	2227	1	POLG HPACV
3	101	100.0	2227	1	POLG HPACV
4	96	95.0	2226	1	POLG HPACV
5	96	95.0	2226	1	POLG HPACV
6	96	95.0	2226	1	POLG HPACV
7	95	95.0	2230	1	POLG HPACV
8	95	94.1	2227	1	POLG HPACV
9	84	83.2	839	1	POLG HPACV
10	75	74.3	341	1	POLG HPACV
11	52.5	52.0	1025	1	DPYD BOVIN
12	45.5	45.0	1025	1	DPYD BOVIN
13	45	44.6	178	1	RLIX ORYSA
14	45	44.6	178	1	RLIX ARATH
15	44	43.6	340	1	RIR2 HSV11
16	44	43.6	340	1	RIR2 HSV1K
17	43	42.6	389	1	DNAJ METMA
18	43	42.6	587	1	MTSL SPRAL
19	42	41.6	178	1	RLIX ARATH
20	41	40.6	289	1	NK26 MOUSE
21	41	40.6	359	1	MCAL CRIGR
22	41	40.6	1040	1	AXOI RAT
23	41	40.6	1234	1	JMJ MOUSE
24	40	39.6	135	1	ZNF7 HAEIN
25	40	39.6	292	1	YOM1 CAEL
26	40	39.6	310	1	MCAL MOUSE
27	40	39.6	312	1	MCAL HUMAN
28	40	39.6	353	1	NK24 HUMAN
29	40	39.6	354	1	NK24 MOUSE
30	40	39.6	371	1	TTT1 CANPA
31	40	39.6	371	1	TTT1 HUMAN
32	40	39.6	372	1	TTT1 MOUSE
33	40	39.6	372	1	TTT1 RAT

34	40	39.6	448	1	DHE4 HELPY
35	40	39.6	523	1	YMR9 YEAST
36	40	39.6	620	1	ARP8 SCHPO
37	40	39.6	659	1	SYM ENTPA
38	40	39.6	1048	1	SRA1 RAT
39	40	39.6	1157	1	SRA4 HUMAN
40	39.5	39.1	1039	1	M2C1 MOUSE
41	39.5	39.1	1040	1	M2C1 RAT
42	39	38.6	178	1	RL1X CASSA
43	39	38.6	204	1	RS4 TRBPA
44	39	38.6	222	1	Y039 RICCN
45	39	38.6	347	1	EL13A SOYBN
46	39	38.6	427	1	TRB1 ARPE
47	39	38.6	462	1	WDR8 MOUSE
48	39	38.6	615	1	CNK RAT
49	39	38.6	631	1	CNK MOUSE
50	39	38.6	646	1	CNK HUMAN
51	39	38.6	673	1	SYM OCEIH
52	39	38.6	808	1	POLG HPACV
53	39	38.6	957	1	IF2 SYNEL
54	39	38.6	1103	1	DPOD RAT
55	39	38.6	1107	1	KCH8 HUMAN
56	39	38.6	1248	1	CYAS RAT
57	39	38.6	1249	1	CYAS MOUSE
58	39	38.6	1251	1	CYAS HUMAN
59	39	38.6	1387	1	RGSC RAT
60	38.5	38.1	585	1	ILV3 YEAST
61	38.5	38.1	881	1	ARPE YEAST
62	38	37.6	49	1	YD03 ARCFU
63	38	37.6	203	1	LOUE VIBPA
64	38	37.6	233	1	RS2 CLOAB
65	38	37.6	317	1	WR15 ARATH
66	38	37.6	398	1	MPK2 CHICK
67	38	37.6	429	1	FXP3 MOUSE
68	38	37.6	465	1	HN4A HUMAN
69	38	37.6	465	1	HN4A MOUSE
70	38	37.6	465	1	HN4A RAT
71	38	37.6	491	1	YA29 SCHPO
72	38	37.6	496	1	MGLA TREPA
73	38	37.6	516	1	ROBG MOUSE
74	38	37.6	752	1	CLPE STRPN
75	38	37.6	1019	1	DLF2 HUMAN
76	38	37.6	1089	1	DLF2 RAT
77	38	37.6	1223	1	KI25 CAEL
78	38	37.6	1266	1	JMJ HUMAN
79	38	37.6	1434	1	VG65 HSV11
80	38	37.6	1787	1	CHD3 CAEL
81	38	37.6	2839	1	NF1 HUMAN
82	37.5	37.1	329	1	GC3 MOUSE
83	37.5	37.1	398	1	GC3M MOUSE
84	37.5	37.1	430	1	CINA MYCTU
85	37	36.6	173	1	YN26 CAEL
86	37	36.6	261	1	RL1 HSV2H
87	37	36.6	293	1	US17 HCMVA
88	37	36.6	301	1	ORT1 HUMAN
89	37	36.6	312	1	O2L2 HUMAN
90	37	36.6	315	1	LIP3 MORSP
91	37	36.6	356	1	MPCP RAT
92	37	36.6	357	1	MPCP MOUSE
93	37	36.6	364	1	MRP AQUAE
94	37	36.6	379	1	TSY1 MOUSE
95	37	36.6	392	1	OR9A DROME
96	37	36.6	413	1	FL1 TOBAN
97	37	36.6	428	1	AFQX XENNE
98	37	36.6	450	1	SR54 EUCBP
99	37	36.6	468	1	PER1 RAT
100	37	36.6	469	1	SYE1 THEMA

ALIGNMENTS

RESULT 1

POLG HPAVC STANDARD; PRT; 852 AA.  
 AC P06442; Q83741; Q83742;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein P2A] (Fragment).  
 OS Hepatitis A virus (strain CR326).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85185648; PubMed=2985793;  
 RA Lineneyer D.L., Menke J.G., Martin-Gallardo A., Hughes J.V.,  
 RA Young A., Mitra S.W.;  
 RT "Molecular cloning and partial sequencing of hepatitis A viral cDNA";  
 RL J. Virol. 54:247-255(1985).  
 CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M10033; AAA5470.1; -;  
 CC PIR; A03904; GNNYHA.  
 CC InterPro; IPR008975; Viral\_cap\_coat.  
 KW Polyprotein; Coat protein; Core protein.  
 FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
 FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
 FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
 FT CHAIN 492 836 COAT PROTEIN VP1 (P1D).  
 FT CHAIN 837 >852 CORE PROTEIN P2A.  
 FT NON TER 852 852  
 SQ SEQUENCE 852 AA; 95563 MW; 73D3ED0AD532820B CRC64;  
 Query Match 100.0%; Score 101; DB 1; Length 852;  
 Best Local Similarity 100.0%; Pred. No. 9.9e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLPPEKMKGLFSQAKISLF 20  
 Db 823 VLPPEKMKGLFSQAKISLF 842  
 RESULT 2  
 ID POLG HPAVC STANDARD; PRT; 2227 AA.  
 AC P06617; P06443; Q81082;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
 OS Hepatitis A virus (strain HM-175).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12098;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Wild type;  
 RX MEDLINE=87061253; PubMed=3023706;  
 RA Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A.,

Baroudy B.M.;  
 "Complete nucleotide sequence of wild-type hepatitis A virus:  
 comparison with different strains of hepatitis A virus and other  
 picornaviruses";  
 J. Virol. 61:50-59(1987).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Attenuated;  
 RX MEDLINE=87175701; PubMed=3031886;  
 RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstein S.M.,  
 RA Purcell R.H.;  
 RT "Complete nucleotide sequence of an attenuated hepatitis A virus:  
 comparison with wild-type virus";  
 Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501(1987).  
 [3]  
 RP SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.  
 RX MEDLINE=85166289; PubMed=2984684;  
 RA Baroudy B.M., Ticehurst J.R., Miele T.A., Maizel J.V. Jr.,  
 RA Purcell R.H., Feinstein S.M.;  
 RT "Sequence analysis of hepatitis A virus cDNA coding for capsid  
 RT proteins and RNA polymerase";  
 Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147(1985).  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA) (N).  
 CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -1- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED  
 CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC -1- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT  
 CC SHOWN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M14114; AAA5475.1; -;  
 CC EMBL; M14707; AAA5465.1; -;  
 CC EMBL; M14707; AAA5466.1; ALT\_INIT.  
 CC EMBL; M16632; AAA5471.1; -;  
 CC PIR; A03905; A03905.  
 CC PIR; A25981; GNNYHM.  
 CC PIR; A94149; GNNYMK.  
 CC PDB; 1HAV; 23-DEC-96.  
 CC MEROPS; C03.005; -;  
 CC InterPro; IPR004004; Calici\_pol\_hel.  
 CC InterPro; IPR009003; Cys\_Ser\_trypsin.  
 CC InterPro; IPR006005; RNA\_helicase.  
 CC InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 CC InterPro; IPR001205; RNA\_pol\_P3D.  
 CC InterPro; IPR007094; RNA\_pol\_P5vir.  
 CC InterPro; IPR008975; Viral\_cap\_coat.  
 CC Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 CC Pfam; PF00910; RNA\_helicase; 1.  
 CC PRINTS; PR00918; CALICIVIRUSN.  
 KW Polyprotein; Coat protein; Core protein; Transferase;  
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.  
 FT CHAIN 1 23  
 FT CHAIN 24 245 COAT PROTEIN VP4 (P1A).  
 FT CHAIN 246 491 COAT PROTEIN VP2 (P1B).  
 FT CHAIN 492 836 COAT PROTEIN VP3 (P1C).  
 FT CHAIN 837 980 COAT PROTEIN VP1 (P1D).  
 FT CHAIN 981 1087 CORE PROTEIN P2A.  
 FT CHAIN 1088 1422 CORE PROTEIN P2B.  
 FT CHAIN 1423 1496 PROBABLE PROTEIN P3A.  
 FT CHAIN 1497 1519 PROBABLE PROTEIN P3B.  
 FT CHAIN 1520 1738 PROBABLE PROTEIN P3C.

FT CHAIN 1739 2227 RNA-DIRECTED POLYMERASE 3D.  
FT VARIANT 77 77 K -> R (IN ATTENUATED STRAIN).  
FT VARIANT 764 764 E -> V (IN ATTENUATED STRAIN).  
FT VARIANT 821 821 N -> S (IN ATTENUATED STRAIN).  
FT VARIANT 1052 1052 A -> V (IN ATTENUATED STRAIN).  
FT VARIANT 1062 1062 G -> A (IN ATTENUATED STRAIN).  
FT VARIANT 1118 1118 K -> M (IN ATTENUATED STRAIN).  
FT VARIANT 1151 1151 E -> K (IN ATTENUATED STRAIN).  
FT VARIANT 1163 1163 F -> S (IN ATTENUATED STRAIN).  
FT VARIANT 1277 1277 V -> I (IN ATTENUATED STRAIN).  
FT VARIANT 1500 1500 H -> Y (IN ATTENUATED STRAIN).  
FT VARIANT 1805 1805 D -> N (IN ATTENUATED STRAIN).  
FT VARIANT 1930 1930 S -> T (IN ATTENUATED STRAIN).  
SQ SEQUENCE 2227 AA; 251506 MW; 01E225E7A5B740A6 CRC64;  
  
Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.7e-08; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VLPPPRKMKGLFSQAKISLF 20  
| | | | | | | | | | | | | | | | | | | | | |  
DB 823 VLPPPRKMKGLFSQAKISLF 842  
  
RESULT 3  
POLG HPVAVL STANDARD; PRT; 2227 AA.  
AC P06441;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain LA).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85190549; PubMed=2986127;  
RA Najarian R., Caput D., Gee W.W., Potter S.J., Renard A.,  
RA Merryweather J., van Nest G., Dina D.;  
RT "Primary structure and gene organization of human hepatitis A virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:2627-2631(1985).  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
(RNA)(N).  
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
each of which is composed of one copy each of proteins VP1, VP2,  
VP3, and VP4.  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; K02990; AAA45472.1; -  
CC PIR; A03903; GNNYHR.  
CC MEROPS; C03.005; -  
CC InterPro; IPR004004; Calici\_pol\_hel.  
CC InterPro; IPR009003; Cys\_ser\_trypsin.  
CC InterPro; IPR000605; RNA\_helicase.  
CC InterPro; IPR007095; RNA\_pol\_DS\_PS.  
CC InterPro; IPR001205; RNA\_pol\_P3D.  
CC InterPro; IPR007094; RNA\_pol\_P3D.  
CC InterPro; IPR008975; Viral\_cap\_coat.  
CC Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
CC Pfam; PF00910; RNA\_helicase; 1.  
  
DR PRINTS; PR00918; CALICVIRUSNS.  
KW Polyprotein; Coat protein; Core protein; Transferase;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
FT CHAIN 492 836 COAT PROTEIN VP1 (P1D).  
FT CHAIN 837 980 CORE PROTEIN P2A.  
FT CHAIN 981 1076 CORE PROTEIN P2B.  
FT CHAIN 1077 1422 CORE PROTEIN P2C.  
FT CHAIN 1423 1484 PROBABLE PROTEIN P3A.  
FT CHAIN 1485 1507 PROBABLE PROTEIN P3B.  
FT CHAIN 1508 1678 PROBABLE PROTEIN P3C.  
FT CHAIN 1679 2227 RNA-DIRECTED POLYMERASE P3D.  
SQ SEQUENCE 2227 AA; 251898 MW; 99A7354B4CD2799C CRC64;  
  
Query Match 100.0%; Score 101; DB 1; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.7e-08; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 VLPPPRKMKGLFSQAKISLF 20  
| | | | | | | | | | | | | | | | | | | | | |  
DB 823 VLPPPRKMKGLFSQAKISLF 842  
  
RESULT 4  
POLG HPVAV2 STANDARD; PRT; 2226 AA.  
AC P26580;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
DE P3D (EC 2.7.7.48)].  
OS Hepatitis A virus (strain 24a).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12094;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91162758; PubMed=1705995;  
RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstein S.M.,  
RA Cromeans T., Jansen R.W.;  
RT "Antigenic and genetic variation in cytopathic hepatitis A virus  
RT variants arising during persistent infection: evidence for genetic  
RT recombination.";  
RL J. Virol. 65:2056-2065(1991).  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
(RNA)(N).  
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
each of which is composed of one copy each of proteins VP1, VP2,  
VP3, and VP4.  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M59810; AAA45468.1; -  
CC MEROPS; C03.005; -  
CC InterPro; IPR004004; Calici\_pol\_hel.  
CC InterPro; IPR009003; Cys\_ser\_trypsin.  
CC InterPro; IPR000605; RNA\_helicase.  
CC InterPro; IPR007095; RNA\_pol\_DS\_PS.  
CC InterPro; IPR001205; RNA\_pol\_P3D.  
CC InterPro; IPR007094; RNA\_pol\_P3D.  
CC InterPro; IPR008975; Viral\_cap\_coat.  
CC Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
CC Pfam; PF00910; RNA\_helicase; 1.

DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR Pfam; PF00910; RNA\_helicase; 1.  
 DR PRINTS; PR00918; CALICVIRUSNS.  
 KW Polypeptide; Coat protein; Core protein; Transferase;  
 FT CHAIN 1 23  
 FT CHAIN 24 245  
 FT CHAIN 246 491  
 FT CHAIN 492 794  
 FT CHAIN 795 900  
 FT CHAIN 901 1087  
 FT CHAIN 1088 1422  
 FT CHAIN 1423 1495  
 FT CHAIN 1496 1518  
 FT CHAIN 1519 1737  
 FT CHAIN 1738 2226  
 SQ SEQUENCE 2226 AA; 251107 MW; 403B4CA80B0BF75 CRC64;  
 Query Match 95.0%; Score 96; DB 1; Length 2226;  
 Best Local Similarity 95.0%; Pred. No. 1.8e-07;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VLPPPRKMKGLFSQAKISLF 20  
 DB 823 VLPPPRKMKGLFSQAKISLF 842  
 RESULT 6  
 POLG\_HPAV8 STANDARD; PRT; 2226 AA.  
 ID P26581;  
 AC P26582;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
 DE Hepatitis A virus (strain 18f).  
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12096;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91162758; PubMed=1705995;  
 RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
 RA Cromeans T., Jansen R.W.;  
 RA "Antigenic and genetic variation in cytopathic hepatitis A virus  
 RT variants arising during persistent infection: evidence for genetic  
 RT recombination.";  
 RL J. Virol. 65:2056-2065(1991).  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M59809; AAA45469.1; --  
 DR MEROPS; C03.005; --  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.

DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR Pfam; PF00910; RNA\_helicase; 1.  
 DR PRINTS; PR00918; CALICVIRUSNS.  
 KW Polypeptide; Coat protein; Core protein; Transferase;  
 FT CHAIN 1 23  
 FT CHAIN 24 245  
 FT CHAIN 246 491  
 FT CHAIN 492 794  
 FT CHAIN 795 900  
 FT CHAIN 901 1087  
 FT CHAIN 1088 1422  
 FT CHAIN 1423 1495  
 FT CHAIN 1496 1518  
 FT CHAIN 1519 1737  
 FT CHAIN 1738 2226  
 SQ SEQUENCE 2226 AA; 251152 MW; 6CD85A91D6B4E2BF CRC64;  
 Query Match 95.0%; Score 96; DB 1; Length 2226;  
 Best Local Similarity 95.0%; Pred. No. 1.8e-07;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VLPPPRKMKGLFSQAKISLF 20  
 DB 823 VLPPPRKMKGLFSQAKISLF 842  
 RESULT 5  
 POLG\_HPAV4 STANDARD; PRT; 2226 AA.  
 ID P26581;  
 AC P26582;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase P3D (EC 2.7.7.48)].  
 DE Hepatitis A virus (strain 43c).  
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12095;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91162758; PubMed=1705995;  
 RA Lemon S.M., Murphy P.C., Shields P.A., Ping L.H., Feinstone S.M.,  
 RA Cromeans T., Jansen R.W.;  
 RA "Antigenic and genetic variation in cytopathic hepatitis A virus  
 RT variants arising during persistent infection: evidence for genetic  
 RT recombination.";  
 RL J. Virol. 65:2056-2065(1991).  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M59809; AAA45469.1; --  
 DR MEROPS; C03.005; --  
 DR InterPro; IPR004004; Calici\_pol\_hel.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR009003; RNA\_helicase.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR001205; RNA\_pol\_P3D.



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; D00924; BAA00766.1; -  
 EMBL; X15461; CAA33490.1; -  
 PIR; A30470; GNNYSA.  
 MEROPS; C03.005; -  
 InterPro; IPR004004; Calici\_pol\_hel.  
 InterPro; IPR009003; Cys\_Ser\_trypsin.  
 InterPro; IPR006005; RNA\_helicase.  
 InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 InterPro; IPR001205; RNA\_pol\_P3D.  
 InterPro; IPR007094; RNA\_pol\_PSVir.  
 InterPro; IPR008975; Viral\_cap\_coat.  
 Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 Pfam; PF00910; RNA\_helicase; 1.  
 PRINTS; PR00918; CALICIVIRUSNS.  
 Polyprotein; Coat protein; Core protein; Transferase;  
 RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
 FT CHAIN 1 27 COAT PROTEIN VP4 (P1A).  
 FT CHAIN 28 249 COAT PROTEIN VP2 (P1B).  
 FT CHAIN 250 495 COAT PROTEIN VP3 (P1C).  
 FT CHAIN 496 795 COAT PROTEIN VP1 (P1D).  
 FT CHAIN 796 984 CORE PROTEIN P2A.  
 FT CHAIN 985 1091 CORE PROTEIN P2B.  
 FT CHAIN 1092 1426 CORE PROTEIN P2C.  
 FT CHAIN 1427 1498 PROBABLE PROTEIN 3A.  
 FT CHAIN 1499 1521 PROBABLE PROTEIN 3B.  
 FT CHAIN 1522 1741 PROBABLE PROTEIN 3C.  
 FT CHAIN 1742 2230 RNA-DIRECTED POLYMERASE 3D.  
 SQ SEQUENCE 2230 AA; 251296 MW; 87B3230E324E1F9 CRC64;

Query Match 95.0%; Score 96; DB 1; Length 2230;  
 Best Local Similarity 90.0%; Pred. No. 1.8e-07;  
 Matches 19; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
 :|||||:|||||:|||||  
 Db 827 ILPPPRKMKGLFSQAKISLF 846

RESULT 8  
 POLG HPVAVM STANDARD; PRT; 2227 AA.  
 ID POLG HPVAVM STANDARD; PRT; 2227 AA.  
 AC P13901; Q81083; Q81084; Q81085; Q81086; Q81087; Q81088; Q81089;  
 AC Q81090; Q81091; Q81092; Q81093;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
 P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
 P3D (EC 2.7.7.48)].  
 DE P3D (EC 2.7.7.48)].  
 OS Hepatitis A virus (strain MB8).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12100;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=88045071; PubMed=2823500;  
 RA Paul A.V., Tada H., der Helm K., Wissel T., Kiehn R., Wimmer E.,  
 RA Deinhardt F.;  
 RA "The entire nucleotide sequence of the genome of human hepatitis A  
 RT virus (isolate MB8).";  
 RL Virus Res. 8:153-171(1987).  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 (RNA)(N).  
 CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -1- MISCELLANEOUS: THIS VIRUS WAS ISOLATED FROM THE STOOL OF A PATIENT

InterPro; IPR006005; RNA\_helicase.  
 InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 InterPro; IPR001205; RNA\_pol\_P3D.  
 InterPro; IPR007094; RNA\_pol\_PSVir.  
 InterPro; IPR008975; Viral\_cap\_coat.  
 Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 Pfam; PF00910; RNA\_helicase; 1.  
 PRINTS; PR00918; CALICIVIRUSNS.  
 Polyprotein; Coat protein; Core protein; Transferase;  
 RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.  
 FT CHAIN 1 23 COAT PROTEIN VP4 (P1A).  
 FT CHAIN 24 245 COAT PROTEIN VP2 (P1B).  
 FT CHAIN 246 491 COAT PROTEIN VP3 (P1C).  
 FT CHAIN 492 794 COAT PROTEIN VP1 (P1D).  
 FT CHAIN 795 900 CORE PROTEIN P2A.  
 FT CHAIN 901 1087 CORE PROTEIN P2B.  
 FT CHAIN 1088 1422 CORE PROTEIN P2C.  
 FT CHAIN 1423 1495 PROBABLE PROTEIN P3A.  
 FT CHAIN 1496 1518 PROBABLE PROTEIN P3B.  
 FT CHAIN 1519 1737 PROBABLE PROTEIN P3C.  
 FT CHAIN 1738 2226 RNA-DIRECTED POLYMERASE P3D.  
 SQ SEQUENCE 2226 AA; 251292 MW; 24964A63396C8D6B CRC64;

Query Match 95.0%; Score 96; DB 1; Length 2226;  
 Best Local Similarity 95.0%; Pred. No. 1.8e-07;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
 :|||||:|||||:|||||  
 Db 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 7  
 POLG HPVAVS  
 ID POLG HPVAVS STANDARD; PRT; 2230 AA.  
 AC P14553;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins  
 P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase  
 P3D (EC 2.7.7.48)].  
 DE P3D (EC 2.7.7.48)].  
 OS Simian hepatitis A virus (strain AGM-27).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12102;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=91311420; PubMed=1649901;  
 RA Tsarev S.A., Emerson S.U., Balayan M.S., Ticehurst J.R.,  
 RA Purcell R.H.;  
 RA "Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome  
 RT structure and growth in cell culture with other HAV strains.";  
 RL J. Gen. Virol. 72:1677-1683(1991).  
 RN [2]  
 RP SEQUENCE OF 1750-2164 FROM N.A.  
 RA MEDLINE=89232168; PubMed=2541023;  
 RA Balayan M.S., Kusov Y.Y., Andjaparidze A.G., Tsarev S.A.,  
 RA Sverdlov E.D., Chizhikov V.E., Blinov V.M., Vasilenko S.K.;  
 RA "Variations in genome fragments coding for RNA polymerase in human  
 RT and simian hepatitis A viruses.";  
 RL FEBS Lett. 247:425-428(1989).  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 (RNA)(N).  
 CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
 CC each of which is composed of one copy each of proteins VP1, VP2,  
 CC VP3, and VP4.  
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC WHO HAD BEEN INFECTED IN THE AFRICAN MEDITERRANEAN REGION.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M20273; AAA45474.1; -.  
CC MEROPS; C03.005; -.  
CC InterPro; IPR004004; Calici\_pol\_hel.  
CC InterPro; IPR009003; Cys\_ser\_trypsin.  
CC InterPro; IPR006005; RNA\_helicase.  
CC InterPro; IPR007095; RNA\_pol\_DS\_PS.  
CC InterPro; IPR001205; RNA\_pol\_P3D.  
CC InterPro; IPR007094; RNA\_pol\_P5vir.  
CC InterPro; IPR008975; Viral\_cap\_coat.  
CC Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
CC Pfam; PF00910; RNA\_helicase; 1.  
CC PRINTS; PR00918; CALICIVIRUSN.  
CC PolyProtein; Coat protein; Core protein; Transferase;  
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
CC CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
CC CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
CC CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
CC CHAIN 492 836 COAT PROTEIN VP1 (PID).  
CC CHAIN 837 980 CORE PROTEIN P2A.  
CC CHAIN 981 1087 CORE PROTEIN P2B.  
CC CHAIN 1088 1422 CORE PROTEIN P2C.  
CC CHAIN 1423 1496 PROBABLE PROTEIN P3A.  
CC CHAIN 1497 1519 PROBABLE PROTEIN P3B.  
CC CHAIN 1520 1738 PROBABLE PROTEIN P3C.  
CC CHAIN 1739 2227 RNA-DIRECTED POLYMERASE 3D.  
CC SEQUENCE 2227 AA; 251435 MW; EC983ED2A7C86349 CRC64;  
CC  
CC Query Match 94.1%; Score 95; DB 1; Length 2227;  
CC Best Local Similarity 95.0%; Pred. No. 2.7e-07;  
CC Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
CC  
CC QY 1 VLPPPRKMKGLFSQAKISLF 20  
CC DB 823 VLPPPRKMKGLFSQAKISLF 842  
CC  
CC RESULT 9  
CC POLG HPVAVT STANDARD; PRT; 839 AA.  
CC AC P31788;  
CC DT 01-JUL-1993 (Rel. 26; Created)  
CC DT 01-JUL-1993 (Rel. 26; Last sequence update)  
CC DT 16-OCT-2001 (Rel. 40; Last annotation update)  
CC DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein  
CC DE P2A] (Fragment).  
CC OS Simian hepatitis A virus (strain CY-145).  
CC OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
CC OC Hepatovirus.  
CC OX NCBI\_TaxID=31707;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RX MEDLINE=91311421; PubMed=1649902;  
CC RA Nainan O.V., Margolis H.S., Robertson B.H., Balayan M., Brinton M.A.;  
CC RT "Sequence analysis of a new hepatitis A virus naturally infecting  
CC RT cynomolgus macaques (Macaca fascicularis).";  
CC RL J. Gen. Virol. 72:1685-1689(1991).  
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M20273; AAA45474.1; -.  
CC MEROPS; C03.005; -.  
CC InterPro; IPR004004; Calici\_pol\_hel.  
CC InterPro; IPR009003; Cys\_ser\_trypsin.  
CC InterPro; IPR006005; RNA\_helicase.  
CC InterPro; IPR007095; RNA\_pol\_DS\_PS.  
CC InterPro; IPR001205; RNA\_pol\_P3D.  
CC InterPro; IPR007094; RNA\_pol\_P5vir.  
CC InterPro; IPR008975; Viral\_cap\_coat.  
CC Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
CC Pfam; PF00910; RNA\_helicase; 1.  
CC PRINTS; PR00918; CALICIVIRUSN.  
CC PolyProtein; Coat protein; Core protein; Transferase;  
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease.  
CC CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
CC CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
CC CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
CC CHAIN 492 836 COAT PROTEIN VP1 (PID).  
CC CHAIN 837 980 CORE PROTEIN P2A.  
CC CHAIN 981 1087 CORE PROTEIN P2B.  
CC CHAIN 1088 1422 CORE PROTEIN P2C.  
CC CHAIN 1423 1496 PROBABLE PROTEIN P3A.  
CC CHAIN 1497 1519 PROBABLE PROTEIN P3B.  
CC CHAIN 1520 1738 PROBABLE PROTEIN P3C.  
CC CHAIN 1739 2227 RNA-DIRECTED POLYMERASE 3D.  
CC SEQUENCE 2227 AA; 251435 MW; EC983ED2A7C86349 CRC64;  
CC  
CC Query Match 94.1%; Score 95; DB 1; Length 2227;  
CC Best Local Similarity 95.0%; Pred. No. 2.7e-07;  
CC Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
CC  
CC QY 1 VLPPPRKMKGLFSQAKISLF 20  
CC DB 823 VLPPPRKMKGLFSQAKISLF 842  
CC  
CC RESULT 9  
CC POLG HPVAVT STANDARD; PRT; 839 AA.  
CC AC P31788;  
CC DT 01-JUL-1993 (Rel. 26; Created)  
CC DT 01-JUL-1993 (Rel. 26; Last sequence update)  
CC DT 16-OCT-2001 (Rel. 40; Last annotation update)  
CC DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core protein  
CC DE P2A] (Fragment).  
CC OS Simian hepatitis A virus (strain CY-145).  
CC OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
CC OC Hepatovirus.  
CC OX NCBI\_TaxID=31707;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RX MEDLINE=91311421; PubMed=1649902;  
CC RA Nainan O.V., Margolis H.S., Robertson B.H., Balayan M., Brinton M.A.;  
CC RT "Sequence analysis of a new hepatitis A virus naturally infecting  
CC RT cynomolgus macaques (Macaca fascicularis).";  
CC RL J. Gen. Virol. 72:1685-1689(1991).  
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M59286; AAA45473.1; -.  
CC PIR; JQ1180; GNNYS2.  
CC InterPro; IPR008975; Viral\_cap\_coat.  
CC PolyProtein; Coat protein; Core protein.  
CC CHAIN 1 23 COAT PROTEIN VP4 (PIA).  
CC CHAIN 24 245 COAT PROTEIN VP2 (PIB).  
CC CHAIN 246 491 COAT PROTEIN VP3 (PIC).  
CC CHAIN 492 ? COAT PROTEIN VP1 (PID).  
CC CHAIN ? >839 CORE PROTEIN P2A.  
CC NON\_TER 839 839  
CC SEQUENCE 839 AA; 93825 MW; 2CACCA4BD1E192DBC CRC64;  
CC  
CC Query Match 83.2%; Score 84; DB 1; Length 839;  
CC Best Local Similarity 83.3%; Pred. No. 6.2e-06;  
CC Matches 15; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
CC  
CC QY 1 VLPPPRKMKGLFSQAKIS 18  
CC DB 822 ILPPPRKLGKGLFSQSKIS 839  
CC  
CC RESULT 10  
CC POLG HPVAVT STANDARD; PRT; 341 AA.  
CC AC P13672;  
CC DT 01-JAN-1990 (Rel. 13; Created)  
CC DT 01-JAN-1990 (Rel. 13; Last sequence update)  
CC DT 10-OCT-2003 (Rel. 42; Last annotation update)  
CC DE Genome polyprotein [Contains: Coat protein VP3 (1C); Coat protein VP1  
CC DE (1D); Core protein P2A] (Fragment).  
CC OS Hepatitis A virus (strain LCD-1).  
CC OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
CC OC Hepatovirus.  
CC OX NCBI\_TaxID=12093;  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RX MEDLINE=89263805; PubMed=2542903;  
CC RA Andonov A.P., Lau P., Chaudhary R.;  
CC RT "Nucleotide sequence of the VP1 gene from a Chinese strain of  
CC RT hepatitis A virus (HAV).";  
CC RL Nucleic Acids Res. 17:3594-3594(1989).  
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral units,  
CC each of which is composed of one copy each of proteins VP1, VP2,  
CC VP3, and VP4.  
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X14666; CAA32794.1; -.  
CC PIR; S04137; S04137.  
CC InterPro; IPR008975; Viral\_cap\_coat.  
CC PolyProtein; Coat protein; Core protein.  
CC NON\_TER 1 1 COAT PROTEIN VP3.  
CC CHAIN <1 1 COAT PROTEIN VP1.  
CC CHAIN 2 340 COAT PROTEIN P2A.  
CC CHAIN 341 >341 CORE PROTEIN P2A.  
CC NON\_TER 341 341  
CC SEQUENCE 341 AA; 38003 MW; 066918289BF136D5 CRC64;  
CC  
CC Query Match 74.3%; Score 75; DB 1; Length 341;

PROPEP	4	1025	DITHYDROPYRIMIDINE DEHYDROGENASE [NADP+]
CHAIN	335	351	NADP (POTENTIAL).
NP_BIND	471	481	FAD (POTENTIAL).
NP_BIND	661	678	URACIL-BINDING (POTENTIAL).
DOMAIN	953	953	IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
METAL	956	956	IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
METAL	959	959	IRON-SULFUR 1 (4FE-4S) (POTENTIAL).
METAL	959	959	IRON-SULFUR 1 (4FE-4S) (POTENTIAL).

uracil and thymidine catabolism and in the pathway leading to the formation of beta-alanine.  
-1- SUBUNIT: Homodimer (By similarity).  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; U20981; AAB40985.1; InterPro; IPR001450; 4Fe4S ferredoxin. InterPro; IPR000759; AdrxnX\_reductase. InterPro; IPR001295; DHO\_dh. InterPro; IPR005720; DHO\_dhl. InterPro; IPR001327; FAD\_Pyr\_redox. InterPro; IPR003009; FMN\_enzyme. InterPro; IPR001100; Pyr\_redox. InterPro; IPR000103; Pyridine\_redox\_2. Pfam; PF01180; DHODEHase; 1. Pfam; PF00037; fer4; 2. Pfam; PF00070; Pyr\_redox; 1. PRINTS; PR00419; ADXROTASE. PRINTS; PR00368; FADPNR. PRINTS; PR00411; PNDROTASEI. PRINTS; PR00469; PNDROTASEII. TIGRFAMs; TIGR01037; pyrd\_subl\_fam; 1. PROSITE; PS00198; 4FE4S\_FERREDOXIN; 1. Oxidoreductase; NADP; Flavoprotein; FAD; FMN; Iron-sulfur; 4Fe-4S. METAL 353 953 IRON-SULFUR 1 (4FE-4S) (POTENTIAL). FT METAL 956 956 IRON-SULFUR 1 (4FE-4S) (POTENTIAL). FT METAL 959 959 IRON-SULFUR 1 (4FE-4S) (POTENTIAL). FT METAL 963 963 IRON-SULFUR 1 (4FE-4S) (POTENTIAL). FT METAL 986 986 IRON-SULFUR 2 (4FE-4S) (POTENTIAL). FT METAL 989 989 IRON-SULFUR 2 (4FE-4S) (POTENTIAL). FT METAL 992 992 IRON-SULFUR 2 (4FE-4S) (POTENTIAL). FT METAL 996 996 IRON-SULFUR 2 (4FE-4S) (POTENTIAL). SQ SEQUENCE 1025 AA; 111696 MW; 5855F93A06C47F4F CRC64;  
Query Match 45.0%; Score 45.5; DB 1; Length 1025;  
Best Local Similarity 61.1%; Pred. No. 17;  
Matches 11; Conservative 2; Mismatches 4; Indels 1; Gaps 1;  
QY 2 LPPPRKMKGLFSQAKISLF 19  
DB 176 LPPPEKMPKPEAYS-AKIAL 192  
RESULT 13  
RLIX ORYSA  
ID RLIX ORYSA STANDARD; PRT; 178 AA.  
AC Q943F3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE 60S ribosomal protein L18a.  
GN RPL18A.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC clone: P0046E05.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration



Db 242 PPRVYGLFRQA 254

RESULT 17

DNAJ METWA

ID DNAJ METWA STANDARD; PRT; 389 AA.

AC P35515;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Chaperone protein dnaJ.

GN DNAJ OR MM2504.

OS Methanosarcina mazei (Methanosarcina frisia).

OC Archaea, Euryarchaeota; Methanomicrobia; Methanosarcinales;

OC Methanosarcinaceae; Methanosarcina.

OX NCBI\_TaxID=2209;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S-6;

RX MEDLINE=93324351; PubMed=8332479;

RA Macario A.J.L., Dugan C.B., Clarens M., Conway de Macario E.;

RT "dnaJ in Archaea.";

RL Nucleic Acids Res. 21:2773-2773 (1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Goel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;

RX MEDLINE=22120827; PubMed=12125824;

RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,

RA Martinek-Arias R., Henne A., Wietzer A., Baeumer S., Jacobi C.,

RA Brueggenmann H., Lienard T., Christmann A., Boencke M., Steckel S.,

RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,

RA Fritz H.-J., Gottschalk G.;

RT "The genome of Methanosarcina mazei: evidence for lateral gene transfer between Bacteria and Archaea.";

RL J. Mol. Microbiol. Biotechnol. 4:453-461 (2002).

CC -!- FUNCTION: Acts as a co-chaperone. Stimulates, jointly with grpE, the ATPase activity of dnaK (By similarity).

CC -!- COFACTOR: Binds 2 zinc ions per monomer (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: Belongs to the dnaJ family.

CC -!- SIMILARITY: Contains 1 J domain.

CC -!- SIMILARITY: Contains 1 CR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X60265; CAA42813.1; -.

CC EMBL; AE013494; AAM32200.1; -.

CC PIR; S41748; S41748.

CC HSP; P25685; IHJG.

CC InterPro; IPR002939; DnaJ C.

CC InterPro; IPR001305; DnaJ CXXCXGKG.

CC InterPro; IPR001623; DnaJ N.

CC InterPro; IPR008971; HSP40 DnaJ pep.

CC InterPro; IPR003095; Hsp\_DnaJ.

CC Pfam; PF00226; DnaJ; 1.

CC Pfam; PF01556; DnaJ C; 1.

CC Pfam; PF00684; DnaJ CXXCXGKG; 1.

CC PRINTS; PR00625; DnaJPROTEIN.

CC SMART; SM00271; DnaJ; 1.

CC PROSITE; PS00636; DnaJ\_1; 1.

CC PROSITE; PS00076; DnaJ\_2; 1.

CC PROSITE; PS00637; DnaJ CXXCXGKG; 1.

CC Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding; Complete proteome.

CC DOMAIN 4 72 J-DOMAIN.

CC DOMAIN 73 110 GLY-RICH.

CC REPEAT 144 151 CXXCXGKG MOTIF.

Query Match 43.6%; Score 44; DB 1; Length 340;

Best Local Similarity 61.5%; Pred. No. 9.3;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQA 15

Db 242 PPRVYGLFRQA 254

RESULT 16

RIR2 HSVIK

ID RIR2 HSVIK STANDARD; PRT; 340 AA.

AC P06474;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Ribonucleoside-diphosphate reductase small chain (EC 1.17.4.1)

DE (Ribonucleoside reductase) (38 kDa subunit).

GN UL40.

OS Herpes simplex virus (type 1 / strain KOS).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Simplexvirus.

OX NCBI\_TaxID=10306;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=83059830; PubMed=6292456;

RA Draper K.G., Frink R.J., Wagner E.K.;

RT "Detailed characterization of an apparently unspliced beta herpes simplex virus type 1 gene mapping in the interior of another.";

RL J. Virol. 43:1123-1128 (1982).

RN [2]

RP REVISIONS.

RA Wagner E.K.;

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Provides the precursors necessary for DNA synthesis.

CC -!- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized thioredoxin + H(2)O = ribonucleoside diphosphate + reduced thioredoxin.

CC -!- COFACTOR: Binds 2 iron ions per subunit (By similarity).

CC -!- PATHWAY: DNA replication pathway; first step.

CC -!- SUBUNIT: Heterodimer of a large and a small chain.

CC -!- SIMILARITY: Belongs to the ribonucleoside diphosphate reductase small chain family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; J02212; AAA66436.1; -.

CC InterPro; IPR000358; Ribonucleotid. red.

CC Pfam; PF00268; ribonuc red sm\_1.

CC PROSITE; PS00368; RIBRED\_SNAUL; 1.

CC Oxidoreductase; DNA replication; Metal-binding; Iron.

CC METAL 94 94 IRON 1 (BY SIMILARITY).

CC METAL 124 124 IRON 1 AND 2 (BY SIMILARITY).

CC METAL 127 127 IRON 1 (BY SIMILARITY).

CC METAL 131 131 BY SIMILARITY.

CC ACT\_SITE 187 187 IRON 2 (BY SIMILARITY).

CC METAL 221 221 IRON 2 (BY SIMILARITY).

CC METAL 224 224 IRON 2 (BY SIMILARITY).

CC SEQUENCE 340 AA; 37966 MW; 921DC04B9D278DE5 CRC64;

Query Match 43.6%; Score 44; DB 1; Length 340;

Best Local Similarity 61.5%; Pred. No. 9.3;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQA 15

FT REPEAT 161 168 CXXCXGKGXG MOTIF.  
 FT REPEAT 187 194 CXXCXGKGXG MOTIF.  
 FT REPEAT 201 208 CXXCXGKGXG MOTIF.  
 FT METAL 144 144 ZINC 1 (BY SIMILARITY).  
 FT METAL 147 147 ZINC 1 (BY SIMILARITY).  
 FT METAL 161 161 ZINC 2 (BY SIMILARITY).  
 FT METAL 164 164 ZINC 2 (BY SIMILARITY).  
 FT METAL 187 187 ZINC 2 (BY SIMILARITY).  
 FT METAL 190 190 ZINC 2 (BY SIMILARITY).  
 FT METAL 201 201 ZINC 1 (BY SIMILARITY).  
 FT METAL 204 204 ZINC 1 (BY SIMILARITY).  
 SQ SEQUENCE 389 AA; 42990 MW; 5B840E379D8C8139 CRC64;  
 Query Match 42.6%; Score 43; DB 1; Length 389;  
 Best Local Similarity 66.7%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 PRKMKGLFSQAK 16  
 |||||:  
 DB 373 PRKMKGLFEVK 384  
 |||||:  
 RESULT 18  
 MSL1\_STRAL STANDARD; PRT; 587 AA.  
 AC Q53609;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Modification methylase Sali (EC 2.1.1.72) (Adenine-specific  
 DE methyltransferase Sali) (M.Sali).  
 GN SALIM.  
 OS Streptomyces albus G.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=1962;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95129852; PubMed=7828868;  
 RA Rodicio M.R., Quinton-Jeger T., Moran L.S., Slatko B.E., Wilson G.G.;  
 RT "Organization and sequence of the Sali restriction-modification  
 RT system".  
 RL Gene 151:167-172(1994).  
 CC -1- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE  
 CC GTCGAC, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND  
 CC PROTECTS THE DNA FROM CLEAVAGE BY THE SALI ENDONUCLEASE.  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-  
 CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U01232; AAA81887.1; -.  
 CC HSP; P14385; 2ADM.  
 CC REBASE; 3491; M.Sali.  
 CC InterPro; IPR002236; N12N6\_mtfase.  
 CC InterPro; IPR002052; N6\_Mtase.  
 CC InterPro; IPR000051; SAM\_bind.  
 CC PRINTS; PR00507; N12N6MTFRASE.  
 CC PROSITE; PS00092; N6\_MTASE; 1.  
 KW Transferase; Methyltransferase; Restriction system.  
 SQ SEQUENCE 587 AA; 64975 MW; 5482F374397FF999 CRC64;  
 Query Match 42.6%; Score 43; DB 1; Length 587;  
 Best Local Similarity 70.0%; Pred. No. 24;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 VLPFPRKMGK 10

Db 334 ILFSPRKMG 343  
 |||||:  
 RESULT 19  
 RL1X\_ARATH STANDARD; PRT; 178 AA.  
 ID RL1X\_ARATH  
 AC P51418; O64699;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE 60S ribosomal protein L18a-1.  
 GN RPL18AA OR AT2G34480 OR T31E10.18.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H.L.,  
 RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Unayam L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,  
 RA Niernan W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,  
 RA Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana".  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=22954850; PubMed=14593172;  
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
 RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Carninci P.,  
 RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,  
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjau M., Hansen N.F.,  
 RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,  
 RA Khan S., Kosemura E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
 RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
 RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
 RA Yuan S., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RT "Empirical analysis of transcriptional activity in the Arabidopsis  
 RT genome".  
 RL Science 302:842-846(2003).  
 RN [3]  
 RP SEQUENCE OF 69-160 FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RP Philipps G., Gigot C.;  
 RA Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.  
 RL -1- SIMILARITY: Belongs to the L18AE family of ribosomal proteins.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AC004077; AAC26708.1; -.  
 CC EMBL; AC004481; AAM14956.1; -.  
 CC EMBL; AY042803; AAK68743.1; -.  
 CC EMBL; Z18039; CAA79087.1; -.  
 CC PIR; T02335; T02335.  
 CC InterPro; IPR002670; Ribosomal\_L18ae.





10-OCT-2003 (Rel. 42, Last annotation update)  
Contactin 2 precursor (Axonin-1) (Axonal glycoprotein TAG-1)  
(transient axonal glycoprotein 1) (TAG-1).  
CWN2 OR TAX1.  
Rattus norvegicus (Rat).  
Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
[1] NCBI\_TaxID=10116;  
SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.  
RC TISSUE=Spinal cord;  
RX MEDLINE=90199890; PubMed=2317872;  
RA Furely A.J., Morton S.B., Manalo D., Karagogeos D., Dodd J.,  
Jessel T.M.;  
RT "The axonal glycoprotein TAG-1 is an immunoglobulin superfamily  
member with neurite outgrowth-promoting activity.";  
RL Cell 61:157-170(1990).  
CC -!- FUNCTION: May play a role in the initial growth and guidance of  
axons. May be involved in cell adhesion.  
CC -!- SUBCELLULAR LOCATION: Attached to the neuronal membrane by a GPI-  
anchor and is also released from neurons.  
CC -!- TISSUE SPECIFICITY: In neural tissues in embryos, and in adult  
brain, spinal cord and cerebellum.  
CC -!- DEVELOPMENTAL STAGE: Transiently expressed on a subset of axons  
in the developing rat nervous system.  
CC -!- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.  
CC -!- SIMILARITY: Contains 4 fibronectin type III domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; M31725; AAA42201.1; -;  
DR PIR; A34695; A34695;  
DR InterPro; IPR008957; FN III-like.  
DR InterPro; IPR003961; FN-III.  
DR InterPro; IPR003962; FNIII subd.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003598; Ig\_c2.  
DR Pfam; PF00041; fn3; 4.  
DR Pfam; PF00047; Ig; 6.  
DR PRINTS; PR00014; FNTYPEIII.  
DR SMART; SM00060; FN3; 4.  
DR SMART; SM00408; IGC2; 5.  
DR PROSITE; PS00835; IG\_LIKE; 6.  
KW Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;  
KW Cell adhesion; Repeat.  
FT SIGNAL 1 30  
FT CHAIN 31 1015  
FT PROPER 1016 1040  
FT DOMAIN 39 130  
FT DOMAIN 135 224  
FT DOMAIN 241 324  
FT DOMAIN 329 413  
FT DOMAIN 419 506  
FT DOMAIN 511 605  
FT DOMAIN 608 614  
FT DOMAIN 613 708  
FT DOMAIN 716 811  
FT DOMAIN 818 910  
FT DOMAIN 911 1005  
FT SITE 796 798  
FT CARBOHYD 78 78  
FT CARBOHYD 200 200  
FT CARBOHYD 206 206  
FT CARBOHYD 463 463  
FT CARBOHYD 479 479  
FT CARBOHYD 500 500  
FT CARBOHYD 527 527

FT CARBOHYD 777 777 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 832 832 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 920 920 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1040 AA; 113042 MW; 6E707EF6614CB4FB CRC64;  
Query Match 40.6%; Score 41; DB 1; Length 1040;  
Best Local Similarity 43.5%; Pred No. 94;  
Matches 10; Conservative 3; Mismatches 6; Indels 4; Gaps 1;  
QY 1 VLPPPRMKG----LFSQAKISL 19  
DB 910 VKPPRRPPGNISWTFSSSLSL 932  
RESULT 23  
ID JMJ\_MOUSE STANDARD; PRT; 1234 AA.  
AC Q62315;  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Jumonji protein.  
GN JMJ.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/Ola;  
RX MEDLINE=95278734; PubMed=7758946;  
RA Takeuchi T., Yamazaki Y., Katoh-Fukui Y., Tsuchiya R., Kondo S.,  
Motoyama J., Higashinakagawa T.;  
RT "Gene trap capture of a novel mouse gene, jumonji, required for  
neural tube formation.";  
RT Genes Dev. 9:1211-1222(1995).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=CS7BL/6; TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RL [3]  
RN SUBCELLULAR LOCATION, AND FUNCTION.  
RX MEDLINE=20374468; PubMed=10913339;  
RA Toyoda M., Kojima M., Takeuchi T.;  
RT "Jumonji is a nuclear protein that participates in the negative  
regulation of cell growth.";  
RL Biochem. Biophys. Res. Commun. 274:332-336(2000).  
RL [4]  
RN SUBCELLULAR LOCATION, AND FUNCTION.  
RX MEDLINE=20269961; PubMed=10807864;  
RA Lee Y., Song A.J., Baker R., Micales B., Conway S.J., Lyons G.E.;  
RT "Jumonji, a nuclear protein that is necessary for normal heart

RT development.",  
 RL Circ. Res. 86:932-938 (2000).  
 CC -1- FUNCTION: Required for neural tube formation. Essential for normal  
 CC heart development and function. Participates in the negative  
 CC regulation of cell proliferation signaling.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Brain.  
 CC -1- SIMILARITY: Contains 1 ARID domain.  
 CC -1- SIMILARITY: Contains 1 JMWJ domain.  
 CC -1- SIMILARITY: Contains 1 JMWJ domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D31967; BAA06736.1; -  
 CC EMBL; BC052444; AAH52444.1; -  
 CC EMBL; BC060695; AAH60695.1; -  
 CC F1; T30254; T30254.  
 CC MGD; MGI:104813; JmJ.  
 CC InterPro; IPR001606; ARID.  
 CC InterPro; IPR003347; TF\_JmJ.  
 CC InterPro; IPR003349; TF\_JmJ.  
 CC InterPro; IPR004198; ZnF\_C5HC2.  
 CC Pfam; PF01368; ARID; 1.  
 CC Pfam; PF02373; JmJC; 1.  
 CC Pfam; PF02375; JmJN; 1.  
 CC Pfam; PF02928; zf-C5HC2; 1.  
 CC SMART; SM00501; BRIGHT; 1.  
 CC SMART; SM00558; JmJC; 1.  
 CC SMART; SM00345; JmJN; 1.  
 CC Developmental protein; Nuclear protein.  
 FT DOMAIN 616 726  
 ARID.  
 SQ SEQUENCE 1234 AA; 137445 MW; 856E172C5E5745B5 CRC64;  
 Query Match 40.6%; Score 41; DB 1; Length 1234;  
 Best Local Similarity 53.8%; Pred. No. 1.1e-02;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 PPRKMKGLFSQAK 16  
 Db 450 PPRKMKGVAGNAE 462  
 RESULT 24  
 ZNTR\_HAEIN STANDARD; PRT; 135 AA.  
 AC P45277;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE HTH-type transcriptional regulator zntr homolog.  
 ZNTR OR H1623.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McElroy K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RL Rd.",  
 RL Science 269:496-512 (1995).  
 CC -1- FUNCTION: Transcriptional regulator (By similarity).  
 CC -1- SIMILARITY: Contains 1 HTH merR-type DNA-binding domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U32835; AAC23268.1; -  
 CC TIGR; H1623; -  
 CC InterPro; IPR000551; HTH\_MerR.  
 CC Pfam; PF00376; merR; 1.  
 CC PRINTS; PR00040; HTHMERR.  
 CC SMART; SM00422; HTH\_MERR; 1.  
 CC PROSITE; PS00552; HTH\_MERR\_1; 1.  
 CC PROSITE; PS00937; HTH\_MERR\_2; 1.  
 CC Transcription regulation; DNA-binding; Complete proteome.  
 KW DOMAIN 1 69 HTH\_MERR-TYPE  
 FT DNA\_BIND 4 23 H-T-H MOTIF (POTENTIAL).  
 SQ SEQUENCE 135 AA; 15636 MW; CE07D84D0BCF9C CRC64;  
 Query Match 39.6%; Score 40; DB 1; Length 135;  
 Best Local Similarity 42.9%; Pred. No. 16;  
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 VLPKPKKGLFSQ 14  
 Db 25 LIPPKRTSGNFRQ 38  
 RESULT 25  
 YOWI\_CABEL STANDARD; PRT; 292 AA.  
 ID YOWI\_CABEL  
 AC P30647;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein ZK643.1 in chromosome III.  
 ZK643.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=92168156; PubMed=1538779;  
 RA Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R.,  
 RA Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A.,  
 RA Craxton M., Durbin R., Berks M., Metzstein M., Hawkins T.,  
 RA Ainscough R., Waterston R.  
 RT "The C. elegans genome sequencing project: a beginning."  
 RL Nature 356:37-41 (1992).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z11126; CAAT7470.1; -  
 CC F1; S23239; S23239.  
 CC WormPep; ZK643.1; CE00440.  
 CC InterPro; IPR000698; Arrestin.

DR InterPro; IPR007110; Ig-like.  
DR Pfam; PF00339; arrestin\_1.  
DR Pfam; PF02752; arrestin\_C; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 292 AA; 32669 MW; 0EDCA0FB82AB6A CRC64;  
  
Query Match 39.6%; Score 40; DB 1; Length 292;  
Best Local Similarity 52.6%; Pred. No. 36;  
Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
  
QY 1 VLPPPRKMKGLFSQAKISL 19  
DB 9 VLAEPKMGAEFFNAKVL 27  
  
RESULT 26  
MCAL\_MOUSE  
ID MCAL\_MOUSE STANDARD; PRT; 310 AA.  
AC P31230; Q60659;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Multisynthetase complex auxiliary component p43 [Contains:  
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (Small  
DE Inducible cytokine subfamily E member 1)].  
GN SCV1 OR EMAP2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95014290; PubMed=7929199;  
RA Kao J., Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L.,  
RA Grikscheit T., Chabot J., Nowyrod R., Greenberg S., Kuang W.J.,  
RA Leung D.W., Hayward J.R., Kiesel W., Heath M., Brett J., Stern D.M.;  
RT "Characterization of a novel tumor-derived cytokine. Endothelial-  
RT monocyte activating polypeptide II";  
RL J. Biol. Chem. 269:25106-25119(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T.I., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.U., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP SEQUENCE OF 145-164 FROM N.A.  
RX MEDLINE=93015897; PubMed=1400342;  
RA Kao J., Ryan J., Brett G., Chen J., Shen H., Fan Y.-G., Godman G.,  
RA Familletti P.C., Wang F., Pan Y.-C.E., Stern D., Claus M.;  
RT "Endothelial monocyte-activating polypeptide II. A novel  
RT tumor-derived polypeptide that activates host-response mechanisms";  
RL J. Biol. Chem. 267:20239-20247(1992).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=94193665; PubMed=7545917;

RA Kao J., Fan Y., Haehnel I., Brett J., Greenberg S., Claus M.,  
RA Kayton M., Houck K., Kiesel W., Seljelid R., Burnier J., Stern D.;  
RT "A peptide derived from the amino terminus of endothelial-monocyte-  
RT activating polypeptide II modulates mononuclear and polymorphonuclear  
RT leukocyte functions, defines an apparently novel cellular interaction  
RT site, and induces an acute inflammatory response.";  
RL J. Biol. Chem. 269:9774-9782(1994).  
CC -I- FUNCTION: ALTERS ENDOTHELIAL AND MONOCYTE FUNCTIONS, INDUCES THE  
CC MIGRATION OF MONOCYTES AND GRANULOCYTES, AND INDUCES AN  
CC INFLAMMATORY RESPONSE IN THE MOUSE FOOTPAD MODEL. EMAP II ELICITS  
CC A PHLOGENIC RESPONSE AND, POTENTIALLY, AUGMENTS THE EFFECTS OF  
CC THE OTHER TUMOR-DERIVED CYTOKINES.  
CC -I- SUBUNIT: Monomer.  
CC -I- SIMILARITY: Contains 1 trna-binding domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC EMBL; U10118; AAA62203.1; -;  
CC EMBL; BC002054; AA02054.1; -;  
CC PIR; A5053; A5053.  
CC MGI; MGI:102774; Scy1.  
CC InterPro; IPR008232; EMAP11.  
CC InterPro; IPR008994; Nucleic acid\_OB.  
CC InterPro; IPR002547; trna\_bind.  
CC Pfam; PF01588; trna\_bind; 1.  
CC PIRSF; PIRSF005381; EMAP11; 1.  
CC PROSITE; PS00886; TRBD; 1.  
CC Protein biosynthesis; RNA-binding; trna-binding; Cytokine.  
CC PROPEP 1 144  
CC CHAIN 145 310  
CC ENDOTHELIAL-MONOCYTE ACTIVATING  
CC POLYPEPTIDE II.  
CC TRNA-BINDING.  
CC DOMAIN 149 250  
CC SEQUENCE 310 AA; 33997 MW; A2F8FF52A3D03A0 CRC64;  
  
Query Match 39.6%; Score 40; DB 1; Length 310;  
Best Local Similarity 53.8%; Pred. No. 38;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 5 PRKMKGLFSQAKI 17  
DB 213 PAKRGVLSQAMV 225  
  
RESULT 27  
MCAL\_HUMAN  
ID MCAL\_HUMAN STANDARD; PRT; 312 AA.  
AC Q12904; Q96CQ9;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Multisynthetase complex auxiliary component p43 [Contains:  
DE Endothelial-monocyte activating polypeptide II (EMAP-II) (Small  
DE Inducible cytokine subfamily E member 1)].  
GN SCV1 OR EMAP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95014290; PubMed=7929199;  
RA Kao J., Houck K., Fan Y., Haehnel I., Libutti S.K., Kayton M.L.,  
RA Grikscheit T., Chabot J., Nowyrod R., Greenberg S., Kuang W.J.,  
RA Leung D.W., Hayward J.R., Kiesel W., Heath M., Brett J., Stern D.M.;  
RT "Characterization of a novel tumor-derived cytokine. Endothelial-  
RT monocyte activating polypeptide II";  
RL J. Biol. Chem. 269:25106-25119(1994).

[2]  
SEQUENCE FROM N.A.  
TISSUE=Pancreas;  
MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences";  
EL Proc Natl Acad Sci U S A. 99:16899-16903(2002).  
CC -1- SIMILARITY: Contains 1 tRNA-binding domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; U10117; AAA62202.1; -;  
DR EMBL; BC014051; AAH14051.1; -;  
DR PDB; 1E7Z; 06-FEB-01.  
DR PDB; 1E7Z; 06-SEP-00.  
DR PDB; 1FLO; 07-FEB-01.  
DR Gene; HGNC:10648; SCYEL.  
DR MIM; 603605; -;  
DR GO; GO:0005615; C:extracellular space; TAS.  
DR GO; GO:0005125; P:cytokine activity; TAS.  
DR GO; GO:0000049; P:tRNA binding; TAS.  
DR GO; GO:0006418; P:amino acid activation; TAS.  
DR GO; GO:0006935; P:chemotaxis; TAS.  
DR GO; GO:0006954; P:inflammatory response; TAS.  
DR InterPro; IPR008232; EMAP11.  
DR InterPro; IPR008994; Nucleic acid OB.  
DR InterPro; IPR002547; tRNA bind.  
DR Pfam; PF01588; tRNA\_bind.1.  
DR PIRSF; PIRSF005381; EMAP11.1.  
DR PROSITE; PS50886; TRBD; 1.  
KW Protein biosynthesis; RNA-binding; tRNA-binding; Cytokine;  
KW 3D-structure.  
FT PROPEP 1 146  
FT CHAIN 147 312  
FT DOMAIN 151 252  
FT CONFLICT 79 79  
FT SEQUENCE 312 AA; 34336 MW; 946310A0216F7587 CRC64;  
Query Match 39.6%; Score 40; DB 1; Length 312;  
Best Local Similarity 53.8%; Pred. No. 39;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 5 PRKMGFLPSQAKI 17  
DB 215 PAHMGVLSQAMV 227  
RESULT 28  
NK24\_HUMAN

DR PROSITE; PS00071; HOMEBOX 2; 1.  
KW DNA-binding; Homeobox; Nuclear protein; Developmental protein.  
FT DNA BIND 189 248 HOMEBOX.  
FT DOMAIN 44 47 POLY-ALA.  
FT DOMAIN 76 91 POLY-ALA.  
FT DOMAIN 181 187 POLY-ALA.  
FT DOMAIN 264 272 POLY-PRO.  
SQ SEQUENCE 353 AA; 36011 MW; B6780C4E4020BED6 CRC64;  
  
Query Match 39.6%; Score 40; DB 1; Length 353;  
Best Local Similarity 57.1%; Pred. No. 44;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 4 PPRKMKGLFSQAKI 17  
Db 188 PRKRRVLFSAQV 201  
  
RESULT 29  
NK24 MOUSE STANDARD; PRT; 354 AA.  
AC Q9EQM3; Q9EQM4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Homeobox protein Nkx-2.4 (Homeobox protein NKX2.4) (Homeobox protein  
DN NK-2 homolog D).  
GN NKX2-4 OR NKX2D.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV, and C57BL/6;  
RX MEDLINE=20279852; PubMed=10818213;  
RA Wang C.-C., Brodnicki T., Copeland N.G., Jenkins N.A., Harvey R.P.;  
RT "Conserved linkage of NK-2 homeobox gene pairs Nkx2-2/2-4 and  
RL Nkx2-1/2-9 in mammals.";  
RN Mamm. Genome 11:466-468(2000).  
RN [2]  
RP PARTIAL SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RX MEDLINE=92153416; PubMed=1346742;  
RA Price M., Lazzaro D., Pohl T., Mattei M.-G., Ruether U., Olivo J.-C.,  
RA Duboule D., Di Lauro R.;  
RT "Regional expression of the homeobox gene Nkx-2.2 in the developing  
RT mammalian forebrain.";  
RL Neuron 8:241-255(1992).  
CC -!- FUNCTION: Probable transcription factor.  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- TISSUE SPECIFICITY: In the embryo it is detected in the posterior  
CC hypothalamus and later in the head. In the adult it is detected  
CC only in testis.  
CC -!- DEVELOPMENTAL STAGE: Expressed in a restricted region of the  
CC posterior hypothalamus from E10. Detected in the head region from  
CC E12.5 to E14.5. Expression is down-regulated by E15.5.  
CC -!- SIMILARITY: Belongs to the NK-2 homeobox family.  
CC -!- SIMILARITY: Contains 1 homeobox domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF202038; AAC35618.1; -.  
DR EMBL; AF202039; AAC35619.1; -.  
DR HSSP; P23441; 1FTT.  
DR MGD; MGI:97349; Nkx2-4.  
DR InterPro; IPR001356; Homeobox.

DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00024; HOMEBOX.  
DR PRODOM; PD00010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX 1; 1.  
DR PROSITE; PS00071; HOMEBOX 2; 1.  
KW DNA-binding; Homeobox; Nuclear protein; Developmental protein.  
FT DNA BIND 188 247 HOMEBOX.  
FT DOMAIN 76 91 POLY-ALA.  
FT DOMAIN 132 136 POLY-ALA.  
FT DOMAIN 263 272 POLY-PRO.  
SQ SEQUENCE 354 AA; 36225 MW; A7A0F450DB15616C CRC64;  
  
Query Match 39.6%; Score 40; DB 1; Length 354;  
Best Local Similarity 57.1%; Pred. No. 44;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 4 PPRKMKGLFSQAKI 17  
Db 187 PRKRRVLFSAQV 200  
  
RESULT 30  
TTF1 CANFA STANDARD; PRT; 371 AA.  
AC P43698;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Thyroid transcription factor 1 (Thyroid nuclear factor 1) (TTF-1)  
DE (Homeobox protein Nkx-2.1).  
GN TTF1 OR TTF1.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96034516; PubMed=7589789;  
RA van Renterghem P.H.G., Dremier S., Vassar G., Christophe J.;  
RT "Study of TTF-1 gene expression in dog thyrocytes in primary  
RT culture.";  
RL Mol. Cell. Endocrinol. 112:83-93(1995).  
CC -!- FUNCTION: Transcription factor that binds and activates the  
CC promoter of thyroid specific genes such as thyroglobulin,  
CC thyroperoxidase, and thyrotropin receptor. Crucial in the  
CC maintenance of the thyroid differentiation phenotype. May play a  
CC role in lung development and surfactant homeostasis.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- TISSUE SPECIFICITY: Thyroid, lung and CNS.  
CC -!- SIMILARITY: Belongs to the NK-2 homeobox family.  
CC -!- SIMILARITY: Contains 1 homeobox domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X77910; CRA54868.1; -.  
DR PIR; I46089; I46089.  
DR HSSP; P23441; 1FTT.  
DR TRANSFAC; T02098; -.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00024; HOMEBOX.  
DR PRODOM; PD00010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX 1; 1.  
DR PROSITE; PS00071; HOMEBOX 2; 1.  
KW Transcription regulation; Activator; Homeobox; DNA-binding;

KW Nuclear protein. 220 HOMOBOX.  
FT DNA BIND 161 220 HOMOBOX.  
FT DOMAIN 234 243 POLY-GLY.  
FT DOMAIN 246 253 POLY-GLN.  
FT DOMAIN 294 303 POLY-ALA.  
SQ SEQUENCE 371 AA; 38539 MW; 3F16CE8E562604D7 CRC64;

Query Match 39.6%; Score 40; DB 1; Length 371;  
Best Local Similarity 57.1%; Pred. No. 46;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 PPRKMKGLFSQAKI 17  
D5 160 PRRKRVLFSAQV 173

Search completed: May 11, 2004, 13:38:05  
Job time : 9.75 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:28:26 ; Search time 32 Seconds  
(without alignments)  
197.199 Million cell updates/sec

Title: US-09-171-432A-42  
Perfect score: 101  
Sequence: 1 VLPFRKMKGLFSQAKISLF 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

- SPTREMBL 25:\*
- 1: sp archaea:\*
  - 2: sp bacteria:\*
  - 3: sp fungi:\*
  - 4: sp human:\*
  - 5: sp invertebrate:\*
  - 6: sp mammal:\*
  - 7: sp mhc:\*
  - 8: sp organelle:\*
  - 9: sp phage:\*
  - 10: sp plant:\*
  - 11: sp rodent:\*
  - 12: sp virus:\*
  - 13: sp vertebrate:\*
  - 14: sp unclassified:\*
  - 15: sp rvirus:\*
  - 16: sp bacteriapi:\*
  - 17: sp archaep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	94	12 Q9ENU9	Q9enu9 hepatitis a
2	101	100.0	94	12 Q9ENV9	Q9env9 hepatitis a
3	101	100.0	94	12 Q9ENV5	Q9env5 hepatitis a
4	101	100.0	94	12 Q9ENV2	Q9env2 hepatitis a
5	101	100.0	94	12 Q9ENV1	Q9env1 hepatitis a
6	101	100.0	94	12 Q9ENV3	Q9env3 hepatitis a
7	101	100.0	94	12 Q9ENU1	Q9enu1 hepatitis a
8	101	100.0	94	12 Q9ENV8	Q9env8 hepatitis a
9	101	100.0	94	12 Q9ENV2	Q9env2 hepatitis a
10	101	100.0	94	12 Q9ENV5	Q9env5 hepatitis a
11	101	100.0	94	12 Q9ENV1	Q9env1 hepatitis a
12	101	100.0	94	12 Q9ENV9	Q9env9 hepatitis a
13	101	100.0	94	12 Q9ENV7	Q9env7 hepatitis a
14	101	100.0	94	12 Q9ENV8	Q9env8 hepatitis a
15	101	100.0	94	12 Q9ENV6	Q9env6 hepatitis a
16	101	100.0	94	12 Q9ENV0	Q9env0 hepatitis a

90 Q9ENT4 Q9ent4 hepatitis a  
91 Q9ENV0 Q9env0 hepatitis a  
92 Q9ENT7 Q9ent7 hepatitis a  
93 Q9ENT6 Q9ent6 hepatitis a  
94 Q86534 Q86534 hepatitis a  
95 Q9dwr5 Q9dwr5 hepatitis a  
96 Q9w7t4 Q9w7t4 hepatitis a  
97 Q71975 Q71975 hepatitis a  
98 Q8b8k2 Q8b8k2 hepatitis a  
99 Q913v2 Q913v2 hepatitis a  
100

ALIGNMENTS

RESULT 1  
Q9ENT9 PRELIMINARY; PRT; 94 AA.  
AC Q9ENT9;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 32;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038298; BAB11836.1; --  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10807 MW; F194CE91BB8C4AFD CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VLPPPRKMKGLFSQAKISLF 20  
Db 74 VLPPPRKMKGLFSQAKISLF 93

RESULT 2  
Q9ENV9 PRELIMINARY; PRT; 94 AA.  
AC Q9ENV9;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 03;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB038278; BAB11816.1; --  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10862 MW; 9AF9EFDAB8C4BAC CRC64;  
Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VLPPPRKMKGLFSQAKISLF 20  
Db 74 VLPPPRKMKGLFSQAKISLF 93

RESULT 3  
Q9ENV5 PRELIMINARY; PRT; 94 AA.  
AC Q9ENV5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 08;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038282; BAB11820.1; --  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VLPPPRKMKGLFSQAKISLF 20  
Db 74 VLPPPRKMKGLFSQAKISLF 93

RESULT 4  
Q9ENV2 PRELIMINARY; PRT; 94 AA.  
AC Q9ENV2;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 28;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";



RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB038295; BAB11833.1; -  
DR PIR; PQ0427; PQ0428.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPPRKMKGLFSQAKISLF 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 74 VLPPPRKMKGLFSQAKISLF 93

## RESULT 5

Q9ENT1 ID Q9ENT1 PRELIMINARY; PRT; 94 AA.

AC Q9ENT1;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Kantou 49;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038306; BAB11844.1; -  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10842 MW; 741850915B86E1A6 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPPRKMKGLFSQAKISLF 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 74 VLPPPRKMKGLFSQAKISLF 93

## RESULT 6

Q9ENV3 ID Q9ENV3 PRELIMINARY; PRT; 94 AA.

AC Q9ENV3;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 12n;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,

RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB038284; BAB11822.1; -  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE8A1A364BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPPRKMKGLFSQAKISLF 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 74 VLPPPRKMKGLFSQAKISLF 93

## RESULT 7

Q9ENU1 ID Q9ENU1 PRELIMINARY; PRT; 94 AA.

AC Q9ENU1;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 29;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038296; BAB11834.1; -  
DR PIR; PQ0427; PQ0427.  
DR PIR; PQ0428; PQ0428.  
DR PIR; PQ0430; PQ0430.  
DR PIR; PQ0431; PQ0431.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPPRKMKGLFSQAKISLF 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 74 VLPPPRKMKGLFSQAKISLF 93

## RESULT 8

Q9ENU8 ID Q9ENU8 PRELIMINARY; PRT; 94 AA.

AC Q9ENU8;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OC NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 22;

```

RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,
RA Yasuoka A., Oka S.;
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB038289; BAB11827.1; -.
DR PIR; PQ0427; PQ0428.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR PIR; PQ0431; PQ0431.
FT NON_TER 1
FT NON_TER 94
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;
Best Local Similarity 100.0%; Pred. No. 2.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20
Db 74 VLPPPRKMKGLFSQAKISLF 93

RESULT 9
Q9ENV2 PRELIMINARY; PRT; 94 AA.
ID Q9ENV2
AC Q9ENV2;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nagasaki 15;
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,
RA Yasuoka A., Oka S.;
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB038285; BAB11823.1; -.
DR PIR; PQ0427; PQ0428.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR PIR; PQ0431; PQ0431.
FT NON_TER 1
FT NON_TER 94
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;
Best Local Similarity 100.0%; Pred. No. 2.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20
Db 74 VLPPPRKMKGLFSQAKISLF 93

RESULT 10
Q9ENV5 PRELIMINARY; PRT; 94 AA.
ID Q9ENV5
AC Q9ENV5;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12092;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nagasaki 25;
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,
RA Yasuoka A., Oka S.;
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB038292; BAB11830.1; -.
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR PIR; PQ0431; PQ0431.
FT NON_TER 1
FT NON_TER 94
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;
Best Local Similarity 100.0%; Pred. No. 2.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20
Db 74 VLPPPRKMKGLFSQAKISLF 93

RESULT 11
Q9ENV1 PRELIMINARY; PRT; 94 AA.
ID Q9ENV1
AC Q9ENV1;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Hepatovirus.
OC NCBI_TaxID=12092;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nagasaki 18;
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,
RA Yasuoka A., Oka S.;
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB038286; BAB11824.1; -.
DR PIR; PQ0427; PQ0427.
DR PIR; PQ0428; PQ0428.
DR PIR; PQ0430; PQ0430.
DR PIR; PQ0431; PQ0431.
DR PIR; PQ0431; PQ0431.
FT NON_TER 1
FT NON_TER 94
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;
Best Local Similarity 100.0%; Pred. No. 2.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20
Db 74 VLPPPRKMKGLFSQAKISLF 93

RESULT 12
Q9ENV9 PRELIMINARY; PRT; 94 AA.
ID Q9ENV9
AC Q9ENV9;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis A virus.

```

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Nagasaki 21;  
 RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
 RA Yasuoka A., Oka S.;  
 RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
 RT infection: Prolonged HAV viremia and mild liver injury.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB038288; BAB11826.1; -  
 DR PIR; PQ0427; PQ0427.  
 DR PIR; PQ0428; PQ0428.  
 DR PIR; PQ0430; PQ0430.  
 DR PIR; PQ0431; PQ0431.  
 FT NON\_TER 1  
 FT NON\_TER 94  
 FT NON\_TER 94  
 SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;  
 Query Match 100.0%; Score 101; DB 12; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLPFRKMKGLFSQAKISLF 20  
 |||||  
 DB 74 VLPFRKMKGLFSQAKISLF 93  
 |||||  
 RESULT 13  
 Q9ENV7  
 ID Q9ENV7 PRELIMINARY; PRT; 94 AA.  
 AC Q9ENV7;  
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Nagasaki 06;  
 RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
 RA Yasuoka A., Oka S.;  
 RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
 RT infection: Prolonged HAV viremia and mild liver injury.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB038280; BAB11818.1; -  
 DR PIR; PQ0427; PQ0427.  
 DR PIR; PQ0428; PQ0428.  
 DR PIR; PQ0430; PQ0430.  
 DR PIR; PQ0431; PQ0431.  
 FT NON\_TER 1  
 FT NON\_TER 94  
 FT NON\_TER 94  
 SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;  
 Query Match 100.0%; Score 101; DB 12; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLPFRKMKGLFSQAKISLF 20  
 |||||  
 DB 74 VLPFRKMKGLFSQAKISLF 93  
 |||||  
 RESULT 14  
 Q9ENV8  
 ID Q9ENV8 PRELIMINARY; PRT; 94 AA.  
 AC Q9ENV8;  
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Nagasaki 04;  
 RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
 RA Yasuoka A., Oka S.;  
 RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
 RT infection: Prolonged HAV viremia and mild liver injury.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB038279; BAB11817.1; -  
 DR PIR; PQ0427; PQ0427.  
 DR PIR; PQ0428; PQ0428.  
 DR PIR; PQ0430; PQ0430.  
 DR PIR; PQ0431; PQ0431.  
 FT NON\_TER 1  
 FT NON\_TER 94  
 FT NON\_TER 94  
 SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;  
 Query Match 100.0%; Score 101; DB 12; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLPFRKMKGLFSQAKISLF 20  
 |||||  
 DB 74 VLPFRKMKGLFSQAKISLF 93  
 |||||  
 RESULT 15  
 Q9ENV6  
 ID Q9ENV6 PRELIMINARY; PRT; 94 AA.  
 AC Q9ENV6;  
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis A virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Hepatovirus.  
 OX NCBI\_TaxID=12092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Nagasaki 07;  
 RA Ida S.;  
 RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
 RT infection: Prolonged HAV viremia and mild liver injury.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB038281; BAB11819.1; -  
 DR PIR; PQ0427; PQ0427.  
 DR PIR; PQ0428; PQ0428.  
 DR PIR; PQ0430; PQ0430.  
 DR PIR; PQ0431; PQ0431.  
 FT NON\_TER 1  
 FT NON\_TER 94  
 FT NON\_TER 94  
 SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;  
 Query Match 100.0%; Score 101; DB 12; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VLPFRKMKGLFSQAKISLF 20  
 |||||  
 DB 74 VLPFRKMKGLFSQAKISLF 93  
 |||||  
 RESULT 16  
 Q9ENW0  
 ID Q9ENW0 PRELIMINARY; PRT; 94 AA.  
 AC Q9ENW0;

```

DT      01-MAR-2001 (TrEMBLrel. 16, Created)
DT      01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Polyprotein (Fragment).
DE      Hepatitis A virus.
OS      Hepatitis A virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC      Hepatovirus.
OX      NCBI_TaxID=12092;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Nagasaki 02;
RA      Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,
RA      Yasuoka A., Oka S.;
RT      "Influence of HIV-1 infection on acute hepatitis A virus (HAV)
RT      infection: Prolonged HAV viraemia and mild liver injury.";
RT      Submitted (F85-3000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AB038277; BABI1815.1; -.
DR      PIR; PQ0427; PQ0427.
DR      PIR; PQ0428; PQ0428.
FT      NON_TER          1
FT      TER              94
FT      NON_TER          94
SQ      SEQUENCE 94 AA; 10903 MW; 9AF8BE85D3D57329C CRC64;

Query Match           100.0%; Score 101; DB 12; Length 94;
Best Local Similarity 100.0%; Pred. No. 2.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

**Qy**            1 VLPPPRKMGIFSQAISLF 20  
             |||||  
**Dd**            74 VLPPPRKMGIFSQAISLF 93

RESULT 17  
Q9ENU6 PRELIMINARY; PRT; 94 AA.

	C9ENCE;	
DT	01-MAR-2001 (TrEMBLrel. 16, Created)	
DT	01-MAR-2001 (TrEMBRel. 16, Last sequence update)	
DT	01-OCT-2003 (TrEMBRel. 25, Last annotation update)	
DE	Polyprotein (Fragment).	
OS	Hepatitis A virus.	
OCC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;	
OCC	Hepatovirus.	
OX	NCBI_TaxID=18292;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Nagasaki 24;	
RA	Iida S., Tachikawa N., Nakajima A., Daiko M., Yano M., Kikuchi Y.,	
RA	Yasuoka A., Oka S.	
RT	"Influence of HIV-1 infection on acute hepatitis A virus (HAV)	
RT	infection: Prolonged HAV viremia and mild liver injury.";	
RL	Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.	
RL	EMBL; AB038291; BAB11829.1; -	
DR	PIR; PQ0427; PQ0427.	
DR	PIR; PQ0428; PQ0428.	
DR	PIR; PQ0430; PQ0430.	
DR	PIR; PQ0431; PQ0431.	
FT	NON_TER 1	
FT	NON_TER 94	
SQ	SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;	

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMGIFSQAISLF 20  
|||  
DB 74 VLPPPRKMGIFSQAISLF 93

RESULT 18  
Q9ENU4  
ID Q9ENU4  
PRELIMINARY: PRT: 94 AA.

RESULT 20  
Q9ENT5 PRELIMINARY; PRT; 94 AA.

ID Q9ENT5  
AC Q9ENT5; 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 23;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038290; BAB11820.1; -.  
DR PIR; P00427; P00428.  
DR PIR; P00428; P00428.  
DR PIR; P00430; P00430.  
DR PIR; P00431; P00431.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10842 MW; 741850915B86E1A6 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||||

Db 74 VLPPPRKMKGLFSQAKISLF 93  
|||||

RESULT 21  
Q9ENU7 PRELIMINARY; PRT; 94 AA.

ID Q9ENU7  
AC Q9ENU7; 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 23;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038290; BAB11820.1; -.  
DR PIR; P00427; P00428.  
DR PIR; P00428; P00428.  
DR PIR; P00430; P00430.  
DR PIR; P00431; P00431.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||||

Db 74 VLPPPRKMKGLFSQAKISLF 93  
|||||

RESULT 22  
Q9ENV4 PRELIMINARY; PRT; 94 AA.

ID Q9ENV4  
AC Q9ENV4; 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 10;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038283; BAB11821.1; -.  
DR PIR; P00427; P00427.  
DR PIR; P00428; P00428.  
DR PIR; P00430; P00430.  
DR PIR; P00431; P00431.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||||

Db 74 VLPPPRKMKGLFSQAKISLF 93  
|||||

RESULT 23  
Q9DWR4 PRELIMINARY; PRT; 115 AA.

ID Q9DWR4  
AC Q9DWR4; 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NOR-26;  
RX MEDLINE=20041342; PubMed=10574047;  
RA Stene-Johansen K., Skaug K., Blystad H.;  
RT "Overvaktning av hepatitt A ved molekylarepidemiologisk  
RT undersøkelse.";  
RL Tidsskr. Nor. Laegeforen. 119:3725-3728 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NOR-26;  
RA Stene-Johansen K.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ299461; CAC14071.2; -.  
DR PIR; P00427; P00427.  
DR PIR; P00428; P00428.  
DR PIR; P00430; P00430.  
DR PIR; P00431; P00431.  
FT NON\_TER 1  
FT CHAIN 1 >64 VP1 PROTEIN.

Db 74 VLPPPRKMKGLFSQAKISLF 93

RESULT 22  
Q9ENV4 PRELIMINARY; PRT; 94 AA.

ID Q9ENV4  
AC Q9ENV4; 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nagasaki 10;  
RA Ida S., Tachikawa N., Nakajima A., Daikoku M., Yano M., Kikuchi Y.,  
RA Yasuoka A., Oka S.;  
RT "Influence of HIV-1 infection on acute hepatitis A virus (HAV)  
RT infection: Prolonged HAV viremia and mild liver injury.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB038283; BAB11821.1; -.  
DR PIR; P00427; P00427.  
DR PIR; P00428; P00428.  
DR PIR; P00430; P00430.  
DR PIR; P00431; P00431.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10876 MW; 9AF8BE91BB8C4BAC CRC64;

Query Match 100.0%; Score 101; DB 12; Length 94;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||||

Db 74 VLPPPRKMKGLFSQAKISLF 93  
|||||

RESULT 23  
Q9DWR4 PRELIMINARY; PRT; 115 AA.

ID Q9DWR4  
AC Q9DWR4; 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE Polypeptide (Fragment).  
OS Hepatitis A virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Hepatovirus.  
OX NCBI\_TaxID=12092;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NOR-26;  
RX MEDLINE=20041342; PubMed=10574047;  
RA Stene-Johansen K., Skaug K., Blystad H.;  
RT "Overvaktning av hepatitt A ved molekylarepidemiologisk  
RT undersøkelse.";  
RL Tidsskr. Nor. Laegeforen. 119:3725-3728 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NOR-26;  
RA Stene-Johansen K.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ299461; CAC14071.2; -.  
DR PIR; P00427; P00427.  
DR PIR; P00428; P00428.  
DR PIR; P00430; P00430.  
DR PIR; P00431; P00431.  
FT NON\_TER 1  
FT CHAIN 1 >64 VP1 PROTEIN.

[illegible]

Db 96 VLPPPRKMKGLFSQAKISLF 115

RESULT 27

QY 071977 PRELIMINARY; PRT; 116 AA.

AC 071977;

DT 01-AUG-1998 (TREMELrel. 07, Created)

DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Polyprotein (Fragment).

OS Hepatitis A virus

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Nor-16;

RX MEDLINE=97442186; PubMed=9298735;

RA Grinde B., Stene-Johansen K., Sharma B., Hoel T., Jensenius M., Skaug K.;

RT "Characterisation of an epidemic of hepatitis A virus involving intravenous drug abusers-infection by needle sharing?";

RL J. Med. Virol. 53:69-75(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Nor-16;

RA Stene-Johansen K., Skaug K., Blystad H., Grinde B.;

RT "A unique hepatitis A virus strain caused an epidemic in Norway associated with intravenous drug abuse.";

RL Scand. J. Infect. Dis. 0:0-0(1998).

DR PIR; P00427; P00428.

DR PIR; P00428; P00428.

FT NON\_TER 1

FT NON\_TER 116

SQ SEQUENCE 116 AA; 13493 MW; 8E26DEB9E7EFBB77 CRC64;

Query Match 100.0%; Score 101; DB 12; Length 116;

Best Local Similarity 100.0%; Pred. No. 3.2e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20

Db 96 VLPPPRKMKGLFSQAKISLF 115

RESULT 28

QY 08B8K6 PRELIMINARY; PRT; 116 AA.

AC 08B8K6;

DT 01-MAR-2003 (TREMELrel. 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)

DE Polyprotein (Fragment).

GN VPI-P2A JUNCTION.

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=IT-AUZ-00;

RA Chironna M.;

RT "Genetic analysis of wild-type strains of HAV recovered from patients with acute hepatitis from southern Italy.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ505621; CAD44276.1; -.

FT NON\_TER 1

FT NON\_TER 116

SQ SEQUENCE 116 AA; 13494 MW; 945B68DF8D8A230C CRC64;

Query Match 100.0%; Score 101; DB 12; Length 116;

Best Local Similarity 100.0%; Pred. No. 3.2e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20

Db 96 VLPPPRKMKGLFSQAKISLF 115

RESULT 29

QY 08B8K5 PRELIMINARY; PRT; 116 AA.

AC 08B8K5;

DT 01-MAR-2003 (TREMELrel. 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)

DE Polyprotein (Fragment).

GN VPI-P2A JUNCTION.

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=IT-DAL-00;

RA Chironna M.;

RT "Genetic analysis of wild-type strains of HAV recovered from patients with acute hepatitis from southern Italy.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ505622; CAD44277.1; -.

FT NON\_TER 1

FT NON\_TER 116

SQ SEQUENCE 116 AA; 13494 MW; 945B68DF8D8A230C CRC64;

Query Match 100.0%; Score 101; DB 12; Length 116;

Best Local Similarity 100.0%; Pred. No. 3.2e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20

Db 95 VLPPPRKMKGLFSQAKISLF 114

RESULT 30

QY 08B8K4 PRELIMINARY; PRT; 116 AA.

AC 08B8K4;

DT 01-MAR-2003 (TREMELrel. 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)

DE Polyprotein (Fragment).

GN VPI-P2A JUNCTION.

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=IT-ILG-00;

RA Chironna M.;

RT "Genetic analysis of wild-type strains of HAV recovered from patients with acute hepatitis from southern Italy.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ505623; CAD44278.1; -.

FT NON\_TER 1

FT NON\_TER 116

SQ SEQUENCE 116 AA; 13494 MW; 945B68DF8D8A230C CRC64;

Query Match 100.0%; Score 101; DB 12; Length 116;

Best Local Similarity 100.0%; Pred. No. 3.2e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20

Db 95 VLPPPRKMKGLFSQAKISLF 114

Best Local Similarity 100.0%; Pred. No. 3.2e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20

Db 95 VLPPPRKMKGLFSQAKISLF 114

RESULT 29

QY 08B8K5 PRELIMINARY; PRT; 116 AA.

AC 08B8K5;

DT 01-MAR-2003 (TREMELrel. 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)

DE Polyprotein (Fragment).

GN VPI-P2A JUNCTION.

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=IT-DAL-00;

RA Chironna M.;

RT "Genetic analysis of wild-type strains of HAV recovered from patients with acute hepatitis from southern Italy.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ505622; CAD44277.1; -.

FT NON\_TER 1

FT NON\_TER 116

SQ SEQUENCE 116 AA; 13494 MW; 945B68DF8D8A230C CRC64;

Query Match 100.0%; Score 101; DB 12; Length 116;

Best Local Similarity 100.0%; Pred. No. 3.2e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20

Db 95 VLPPPRKMKGLFSQAKISLF 114

RESULT 30

QY 08B8K4 PRELIMINARY; PRT; 116 AA.

AC 08B8K4;

DT 01-MAR-2003 (TREMELrel. 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)

DE Polyprotein (Fragment).

GN VPI-P2A JUNCTION.

OS Hepatitis A virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Hepatovirus.

OX NCBI\_TaxID=12092;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=IT-ILG-00;

RA Chironna M.;

RT "Genetic analysis of wild-type strains of HAV recovered from patients with acute hepatitis from southern Italy.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ505623; CAD44278.1; -.

FT NON\_TER 1

FT NON\_TER 116

SQ SEQUENCE 116 AA; 13494 MW; 945B68DF8D8A230C CRC64;

Query Match 100.0%; Score 101; DB 12; Length 116;

Best Local Similarity 100.0%; Pred. No. 3.2e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20

Db 95 VLPPPRKMKGLFSQAKISLF 114

Db 95 VLPPRRKMGIFSQAKISLF 114

Search completed: May 11, 2004, 13:40:25  
Job time : 33 secs





## ALIGNMENTS

RESULT 1  
5516630-2  
; Patent No. 5516630  
; APPLICANT: TICEHURST, JOHN R.; BALTIMORE, DAVID; FEINSTONE,  
; STEPHEN M.; PURCELL, ROBERT H.; RACANELLO, VINCENT R.;  
; BAROUDY, BAIGIE M.  
; TITLE OF INVENTION: METHODS OF DETECTING HEPATITIS A VIRUS  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION NUMBER: US/07/789,262  
; FILING DATE: 06-NOV-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 256,135  
; FILING DATE: 06-OCT-1988  
; APPLICATION NUMBER: 654,942  
; FILING DATE: 27-SEP-1984  
; APPLICATION NUMBER: 537,911  
; FILING DATE: 30-SEP-1983  
; SEQ ID NO: 2:  
; LENGTH: 1091  
5516630-2

Query Match 100.0%; Score 101; DB 6; Length 1091;  
Best Local Similarity 100.0%; Pred. No. 1.2e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||  
Db 1060 VLPPPRKMKGLFSQAKISLF 1079

RESULT 2  
US-08-475-886-2  
; Sequence 2, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-08-475-886-2

Query Match 100.0%; Score 101; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||  
Db 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 3  
US-08-475-886-4  
; Sequence 4, Application US/08475886A

; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated HAV (Pass 35), strain HM-175  
US-08-475-886-4

Query Match 100.0%; Score 101; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||  
Db 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 4  
US-08-475-886-6  
; Sequence 6, Application US/08475886A  
; Patent No. 6113912  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H  
; APPLICANT: D'HONDT, ERIC  
; TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
; FILE REFERENCE: 20264262US2  
; CURRENT APPLICATION NUMBER: US/08/475,886A  
; CURRENT FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 07/947,338  
; EARLIER FILING DATE: 1992-09-18  
; EARLIER APPLICATION NUMBER: 08/397,232  
; EARLIER FILING DATE: 1995-03-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2227  
; TYPE: PRT  
; ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-08-475-886-6

Query Match 100.0%; Score 101; DB 3; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
|||  
Db 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 5  
US-08-397-232-2  
; Sequence 2, Application US/08397232A  
; Patent No. 6180110  
; GENERAL INFORMATION:  
; APPLICANT: FUNKHOUSER, ANN W  
; APPLICANT: EMERSON, SUZANNE U  
; APPLICANT: PURCELL, ROBERT H



```
Query Match      100.0%; Score 101; DB 4; Length 2227;
Best Local Similarity 100.0%; Pred. No. 2.6e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0;
```

Query Match 100.0%; Score 101; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMGKGLFSQAKISLF 20  
|||  
db 823 VLPPPRKMGKGLFSQAKISLF 842

GENERAL INFORMATION:  
APPLICANT: NAINAN, OMANA V.  
APPLICANT: MARGOLIS, HAROLD S.  
APPLICANT: ROBERTSON, BETTY H.  
APPLICANT: BRINTON, MARGO H.  
APPLICANT: EBERT, JAMES W.  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1615 L Street N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,016  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,828  
FILING DATE: 03-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: SCOTT, WATSON I.  
REGISTRATION NUMBER: 26,581  
REFERENCE/DOCKET NUMBER: WTS/5683/83834/SRL  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 839 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-087-016-2

Query Match 83.2%; Score 84; DB 1; Length 839;  
Best Local Similarity 83.3%; Pred. No. 6.6e-06;  
Matches 15; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPPRKMKGLFSQAKISLF 18  
Db 822 ILPPPRKLGKGLFSQSKIS 839

RESULT 16  
US-08-304-309-2  
Sequence 2, Application US/08304309  
Patent No. 5856454  
GENERAL INFORMATION:  
APPLICANT: GONZALEZ, Frank J.  
APPLICANT: FERNANDEZ-SALGUERO, Pedro  
TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN  
TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: Steuart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

RESULT 13  
US-10-135-988-4  
Sequence 4, Application US/10135988  
Patent No. 6680060  
GENERAL INFORMATION:  
APPLICANT: FUNKHOUSER, ANN W  
APPLICANT: EMERSON, SUZANNE U  
APPLICANT: PURCELL, ROBERT H  
APPLICANT: D'HONDT, ERIC  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
FILE REFERENCE: 20264262US3  
CURRENT APPLICATION NUMBER: US/10/135,988  
CURRENT FILING DATE: 2002-04-29  
PRIOR APPLICATION NUMBER: 07/947,338  
PRIOR FILING DATE: 1992-09-18  
PRIOR APPLICATION NUMBER: 08/397,232  
PRIOR FILING DATE: 1995-03-10  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 4  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: Attenuated HAV (Pass 35), strain HM-175  
US-10-135-988-4

Query Match 100.0%; Score 101; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPPRKMKGLFSQAKISLF 20  
Db 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 14  
US-10-135-988-6  
Sequence 6, Application US/10135988  
Patent No. 6680060  
GENERAL INFORMATION:  
APPLICANT: FUNKHOUSER, ANN W  
APPLICANT: EMERSON, SUZANNE U  
APPLICANT: PURCELL, ROBERT H  
APPLICANT: D'HONDT, ERIC  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
FILE REFERENCE: 20264262US3  
CURRENT APPLICATION NUMBER: US/10/135,988  
CURRENT FILING DATE: 2002-04-29  
PRIOR APPLICATION NUMBER: 07/947,338  
PRIOR FILING DATE: 1992-09-18  
PRIOR APPLICATION NUMBER: 08/397,232  
PRIOR FILING DATE: 1995-03-10  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 6  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: Attenuated (4380) HAV, strain HM-175  
US-10-135-988-6

Query Match 100.0%; Score 101; DB 4; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 2.6e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLPPPRKMKGLFSQAKISLF 20  
Db 823 VLPPPRKMKGLFSQAKISLF 842

RESULT 15  
US-08-087-016-2  
Sequence 2, Application US/08087016  
Patent No. 5430135

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/304,309  
APPLICATION NUMBER: US/08/304,309  
FILING DATE: 09-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Timothy L.  
REGISTRATION NUMBER: 35,367  
REFERENCE/DOCKET NUMBER: 15280-210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-304-309-2

Query Match 52.0%; Score 52.5; DB 2; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 1.7;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

Qy 2 LPPPRKMKGLFSQAKISLF 20  
Db 176 LPPPEKMKSEAYS-AKIALF 193

RESULT 17  
US-08-991-942-2  
Sequence 2, Application US/08991942  
Patent No. 6015673  
GENERAL INFORMATION:  
APPLICANT: GONZALEZ, Frank J.  
APPLICANT: FERNANDEZ-SALGUERO, Pedro  
TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN  
TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: Steuart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/991,942  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/304,309  
FILING DATE: 09-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Timothy L.  
REGISTRATION NUMBER: 35,367  
REFERENCE/DOCKET NUMBER: 15280-210  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-991-942-2

Query Match 52.0%; Score 52.5; DB 3; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 1.7;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;  
Qy 2 LPPPRKMKGLFSQAKISLF 20  
Db 176 LPPPEKMKSEAYS-AKIALF 193  
RESULT 18  
US-09-138-103-2  
Sequence 2, Application US/09138103A  
Patent No. 6232448  
GENERAL INFORMATION:  
APPLICANT: Yoshikubo, Takashi  
APPLICANT: Hasegawa, Masami  
TITLE OF INVENTION: Dihydropyrimidine Dehydrogenase  
TITLE OF INVENTION: Dihydropyrimidine Dehydrogenase  
FILE REFERENCE: 09/138,103 Yoshikubo, et al.  
CURRENT APPLICATION NUMBER: US/09/138,103A  
CURRENT FILING DATE: 1998-08-21  
EARLIER APPLICATION NUMBER: 97114630.3  
EARLIER FILING DATE: 1997-08-22  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1025  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-138-103-2

Query Match 52.0%; Score 52.5; DB 3; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 1.7;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

Qy 2 LPPPRKMKGLFSQAKISLF 20  
Db 176 LPPPEKMKSEAYS-AKIALF 193

RESULT 19  
PCT-US95-04567-4  
Sequence 4, Application PC/TUS9504567  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE  
TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: PatentIn Release #1.0, Version  
SOFTWARE: #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04567  
FILING DATE: CONCURRENTLY HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,357  
FILING DATE: 13-APR-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B.  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: UOAB025P--

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04567-4

Query Match 52.0%; Score 52.5; DB 5; Length 1025;  
Best Local Similarity 63.2%; Pred. No. 1.7;  
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 2 LPPPRKMKGLFSQAKISLF 20  
DB 176 LPPPRKMKGLFSQAKISLF 193

RESULT 20  
PCT-US95-04567-2  
Sequence 2, Application PC/TUS9504567  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE  
TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04567  
FILING DATE: CONCURRENTLY HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/227,357  
FILING DATE: 13-APR-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B.  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: UOAB025P--  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1025 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04567-2

Query Match 45.0%; Score 45.5; DB 5; Length 1025;  
Best Local Similarity 61.1%; Pred. No. 26;  
Matches 11; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 2 LPPPRKMKGLFSQAKISLF 19  
DB 176 LPPPRKMKGLFSQAKISLF 192

RESULT 21  
US-07-800-364B-4  
Sequence 4, Application US/07800364B  
Patent No. 5688678  
GENERAL INFORMATION:  
APPLICANT: Hewick, Rodney M.  
APPLICANT: Wang, Jack H.  
APPLICANT: Wozney, John H.  
APPLICANT: Celeste, Anthony J.  
TITLE OF INVENTION: Bone and Cartilage Inductive Proteins  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Legal Affairs, Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/800,364B  
FILING DATE: 26-NOV-1991  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Kapinos, Ellen J.  
REGISTRATION NUMBER: 32,245  
REFERENCE/DOCKET NUMBER: GI 5182A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-876-1170  
TELEFAX: 617-876-5851  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Bos taurus  
TISSUE TYPE: Bone  
US-07-800-364B-4

Query Match 44.6%; Score 45; DB 1; Length 23;  
Best Local Similarity 63.6%; Pred. No. 0.51;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LPPPRKMKGLF 12  
DB 4 LPPPRKMKGLF 14

RESULT 22  
PCT-US91-03388-4  
Sequence 4, Application PC/TUS9103388  
GENERAL INFORMATION:  
APPLICANT: Hewick, Rodney M.  
APPLICANT: Wang, Jack H.  
TITLE OF INVENTION: Bone and Cartilage Inductive Proteins  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Legal Affairs, Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/03388
; FILING DATE: 19910515
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Kapinos, Ellen J.
; REGISTRATION NUMBER: 32,245
; REFERENCE/DOCKET NUMBER: G15182X-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-876-1170
; TELEFAX: 617-876-5851
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; TISSUE TYPE: Bone
; PCT-US91-03388-4

Query Match 44.6%; Score 45; DB 5; Length 23;
Best Local Similarity 63.6%; Pred. No. 0.51;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LPPPRKMKGLF 12
| | | | |
DB 4 LPPPNKLPGLF 14

RESULT 23
US-09-134-001C-4438
; Sequence 4438, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4438
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4438

Query Match 43.6%; Score 44; DB 4; Length 454;
Best Local Similarity 53.3%; Pred. No. 19;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 LPPPRKMKGLFSQAK 16
| | | | |
DB 179 IPFPDKYKGMFEQAK 193

RESULT 24
US-09-205-258-855
; Sequence 855, Application US/09205258
; Patent No. 6525174

```

GENERAL INFORMATION:

```

; APPLICANT: Young et al
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: PZ007P1
; CURRENT APPLICATION NUMBER: US/09/205,258
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,893
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,900
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,901
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,892
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,915
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,019
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,972
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,916
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,373
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,875
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,374
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,917
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,949
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,883
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962

```



; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,963  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,877  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,878  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/070,923  
; EARLIER FILING DATE: 1997-12-18  
; EARLIER APPLICATION NUMBER: 60/092,921  
; EARLIER FILING DATE: 1998-07-15  
; EARLIER APPLICATION NUMBER: 60/094,657  
; EARLIER FILING DATE: 1998-07-30  
; NUMBER OF SEQ ID NOS: 1227  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 855  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-205-258-855

Query Match 43.1%; Score 43.5; DB 4; Length 97;  
Best Local Similarity 47.4%; Pred. No. 4.3;  
Matches 9; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Qy 2 LPPrKMKGLFSQAKISLF 20  
Db 52 VPPLKMPGVLA-IRLSLF 69

RESULT 25  
US-09-252-991A-30726  
; Sequence 30726, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30726  
; LENGTH: 345  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30726

Query Match 41.6%; Score 42; DB 4; Length 345;  
Best Local Similarity 50.0%; Pred. No. 31;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 4 PPRKMKGLFSQAKISL 19  
Db 12 PARETRGLTSQKRVSL 27

RESULT 26  
US-09-252-991A-30833  
; Sequence 30833, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30833  
; LENGTH: 1197  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; NAME/KEY: UNSURE  
; LOCATION: (1028)  
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.  
US-09-252-991A-30833

Query Match 41.6%; Score 42; DB 4; Length 1197;  
Best Local Similarity 62.5%; Pred. No. 1.2e+02;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VLPPRRKMKGLFSQAK 16  
Db 1012 VPPRRARKLRQAK 1027

RESULT 27  
US-09-252-991A-23794  
; Sequence 23794, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23794  
; LENGTH: 330  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23794

Query Match 41.1%; Score 41.5; DB 4; Length 330;  
Best Local Similarity 50.0%; Pred. No. 36;  
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Qy 2 LPPrKMKGLFSQ-AKISLF 20  
Db 128 IPASSRFKGTFSQAEIQLF 147

RESULT 28  
US-09-621-976-7008  
; Sequence 7008, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTS and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 7008  
; LENGTH: 51  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-621-976-7008



; MOLECULE TYPE: protein  
US-08-444-733-106

Query Match 40.6%; Score 41; DB 2; Length 108;  
Best Local Similarity 46.7%; Pred. No. 13;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQAKI 17  
||| ||| : :  
Db 57 PPRVPAKGVPEAPL 71

Search completed: May 11, 2004, 13:42:23  
Job time : 15.5 secs

OM protein - protein search, using sw model

Run on: May 11, 2004, 13:37:37 ; Search time 34.25 Seconds  
(without alignments)  
162.083 Million cell updates/sec

Title: US-09-171-432A-42

Perfect score: 101

Sequence: 1 VLPPPRKMKGLFSQAKISLF 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US05\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	101	100.0	20	10	US-09-171-432A-42
2	101	100.0	352	14	US-10-272-459-45
3	101	100.0	980	14	US-10-272-459-41
4	101	100.0	2227	9	US-09-929-955-12
5	101	100.0	2227	13	US-10-104-966-12
6	101	100.0	2227	13	US-10-135-988-2
7	101	100.0	2227	13	US-10-135-988-4
8	101	100.0	2227	13	US-10-135-988-6
9	70	69.3	836	14	US-10-272-459-40
10	65	64.4	25	10	US-09-171-432A-47
11	52.5	52.0	1025	9	US-09-854-886-2
12	51	50.5	184	12	US-10-424-599-242191
13	48	47.5	187	12	US-10-424-599-264591
14	48	47.5	524	15	US-10-320-797-3221
15	47	46.5	650	12	US-10-424-599-280225

16	46	45.5	73	12	US-10-424-599-196137	Sequence 196137,
17	46	45.5	158	12	US-10-424-599-265499	Sequence 265499,
18	45	44.6	171	15	US-10-094-749-2193	Sequence 2193, Ap
19	45	44.6	2627	12	US-10-424-599-256710	Sequence 256710,
20	45	44.6	2834	12	US-10-424-599-2056711	Sequence 205671,
21	44	43.6	151	12	US-10-424-599-200960	Sequence 200960,
22	44	43.6	178	12	US-10-424-599-198250	Sequence 198250,
23	44	43.6	408	12	US-10-087-584-76	Sequence 76, Appl
24	44	43.6	408	12	US-10-218-779-76	Sequence 76, Appl
25	44	43.6	457	12	US-10-425-114-43634	Sequence 43634, A
26	44	43.6	490	12	US-10-425-114-43063	Sequence 43063, A
27	44	43.6	522	15	US-10-104-047-3665	Sequence 3665, Ap
28	44	43.6	673	14	US-10-176-306-2	Sequence 2, Appl
29	43.5	43.1	62	12	US-10-424-599-212719	Sequence 212719,
30	43.5	43.1	97	10	US-09-933-767-855	Sequence 855, App
31	43.5	43.1	97	12	US-10-004-860-855	Sequence 855, App
32	43.5	43.1	97	14	US-10-023-282-855	Sequence 855, App
33	43	42.6	178	12	US-10-424-599-192572	Sequence 192572,
34	43	42.6	178	12	US-10-424-599-192574	Sequence 192574,
35	42.5	42.1	80	12	US-10-424-599-261181	Sequence 261181,
36	42.5	42.1	103	12	US-10-424-599-284569	Sequence 284569,
37	42	41.6	20	10	US-09-171-432A-43	Sequence 43, Appl
38	42	41.6	53	12	US-10-424-599-263190	Sequence 263190,
39	42	41.6	78	12	US-10-424-599-165621	Sequence 165621,
40	42	41.6	79	14	US-10-106-598-4310	Sequence 4310, Ap
41	42	41.6	83	12	US-10-424-599-153362	Sequence 153362,
42	42	41.6	113	11	US-09-854-408A-4034	Sequence 4034, Ap
43	42	41.6	178	12	US-10-424-599-151916	Sequence 151916,
44	42	41.6	255	12	US-10-424-599-174296	Sequence 174296,
45	42	41.6	270	12	US-10-282-122A-67337	Sequence 67337, A
46	42	41.6	449	14	US-10-156-761-13588	Sequence 13588, A
47	42	41.6	669	12	US-10-282-122A-72423	Sequence 72423, A
48	42	41.6	847	12	US-10-282-511-8	Sequence 8, Appl
49	42	41.6	2053	13	US-10-017-216-2	Sequence 2, Appl
50	42	41.6	2053	14	US-10-325-430-12	Sequence 12, Appl
51	41.5	41.1	49	12	US-10-424-599-153979	Sequence 153979,
52	41.5	41.1	76	12	US-10-424-599-185254	Sequence 185254,
53	41	40.6	60	12	US-10-424-599-281008	Sequence 281008,
54	41	40.6	86	12	US-10-424-599-278781	Sequence 278781,
55	41	40.6	113	15	US-10-131-410-89	Sequence 89, Appl
56	41	40.6	156	12	US-10-424-599-15918	Sequence 15918,
57	41	40.6	178	12	US-10-424-599-151918	Sequence 151918,
58	41	40.6	194	12	US-10-424-599-221615	Sequence 221615,
59	41	40.6	306	12	US-10-424-599-179002	Sequence 179002,
60	41	40.6	2842	14	US-10-224-399A-3479	Sequence 3479, Ap
61	41	40.6	2905	8	US-08-424-550B-401	Sequence 401, App
62	41	40.6	2910	10	US-09-828-498-2	Sequence 2, Appl
63	40.5	40.1	86	12	US-10-424-599-165143	Sequence 165143,
64	40.5	40.1	105	12	US-10-424-599-259408	Sequence 259408,
65	40	39.6	64	12	US-10-424-599-163834	Sequence 163834,
66	40	39.6	95	12	US-10-424-599-193808	Sequence 193808,
67	40	39.6	101	15	US-10-264-049-4089	Sequence 4089, Ap
68	40	39.6	113	12	US-10-424-599-147825	Sequence 147825,
69	40	39.6	137	12	US-10-424-599-220719	Sequence 220719,
70	40	39.6	166	9	US-09-851-026-11	Sequence 11, Appl
71	40	39.6	166	10	US-09-930-169-3	Sequence 3, Appl
72	40	39.6	178	12	US-10-424-599-281376	Sequence 281376,
73	40	39.6	182	12	US-10-424-599-194866	Sequence 194866,
74	40	39.6	183	12	US-10-424-599-181763	Sequence 181763,
75	40	39.6	272	10	US-09-934-455-140	Sequence 140, App
76	40	39.6	272	15	US-10-325-068-200	Sequence 200, App
77	40	39.6	272	15	US-10-374-780A-2428	Sequence 2428, Ap
78	40	39.6	310	9	US-09-851-026-35	Sequence 35, Appl
79	40	39.6	312	9	US-09-851-026-36	Sequence 36, Appl
80	40	39.6	312	9	US-09-947-971-2	Sequence 2, Appl
81	40	39.6	320	12	US-09-925-298-472	Sequence 472, App
82	40	39.6	320	14	US-10-102-806-472	Sequence 472, App
83	40	39.6	354	12	US-10-425-114-65632	Sequence 65632, A
84	40	39.6	371	9	US-09-320-337-76	Sequence 76, Appl
85	40	39.6	385	12	US-10-424-599-237737	Sequence 237737,
86	40	39.6	401	15	US-10-236-031B-26	Sequence 26, Appl
87	40	39.6	406	12	US-10-282-122A-73403	Sequence 73403, A
88	40	39.6	448	12	US-10-282-122A-58700	Sequence 58700, A

89 40 39.6 453 12 US-10-282-122A-52293 Sequence 52293, A  
90 40 39.6 466 12 US-10-435-114-37460 Sequence 37460, A  
91 40 39.6 524 12 US-10-072-012-230 Sequence 230, App  
92 40 39.6 562 12 US-10-425-114-70255 Sequence 70255, A  
93 40 39.6 562 15 US-10-369-493-21428 Sequence 21428, A  
94 40 39.6 666 9 US-09-815-242-4933 Sequence 4933, Ap  
95 40 39.6 669 12 US-10-282-122A-57809 Sequence 57809, A  
96 40 39.6 677 9 US-09-815-242-10663 Sequence 10663, A  
97 40 39.6 677 12 US-10-282-122A-42475 Sequence 42475, A  
98 40 39.6 679 12 US-10-282-122A-48489 Sequence 48489, A  
99 40 39.6 851 10 US-09-991-936-1915 Sequence 1915, Ap  
100 40 39.6 2623 12 US-10-412-699B-1519 Sequence 1519, Ap

#### ALIGNMENTS

RESULT 1  
US-09-171-432A-42  
; Sequence 42, Application US/09171432A  
; Publication No. US20030187194A1  
; GENERAL INFORMATION:  
; APPLICANT: Fields, Howard A.  
; APPLICANT: Khudyakov, Yuri E.  
; TITLE OF INVENTION: Antigenically Reactive Regions of the  
; TITLE OF INVENTION: Hepatitis A Virus Polyprotein  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick Stockton LLP  
; STREET: 3424 Peachtree Road, N.E.  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30326  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/171,432A  
; FILING DATE: 23-NOV-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Greene, Jamie L.  
; REGISTRATION NUMBER: 32,467  
; REFERENCE/DOCKET NUMBER: 03063-0231US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 949-2400  
; TELEFAX: (404) 949-2499  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..20  
; OTHER INFORMATION: /label=YK-1318

US-09-171-432A-42  
Query Match 100.0%; Score 101; DB 10; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VLPPPRKMKGLFSQAKISLF 20  
|||  
DB 1 VLPPPRKMKGLFSQAKISLF 20  
|||

RESULT 2  
US-10-272-459-45  
; Sequence 45, Application US/10272459  
; Publication No. US20030124517A1  
; GENERAL INFORMATION:  
; APPLICANT: PICHUANES, Sergio  
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; FILE REFERENCE: PPI7955.002 / 2301-17955  
; CURRENT APPLICATION NUMBER: US/10/272,459  
; CURRENT FILING DATE: 2002-10-15  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 352  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein  
; OTHER INFORMATION: of 38.8 kDa  
US-10-272-459-45

Query Match 100.0%; Score 101; DB 14; Length 352;  
Best Local Similarity 100.0%; Pred. No. 2.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VLPPPRKMKGLFSQAKISLF 20  
|||  
DB 195 VLPPPRKMKGLFSQAKISLF 214  
|||

RESULT 3  
US-10-272-459-41  
; Sequence 41, Application US/10272459  
; Publication No. US20030124517A1  
; GENERAL INFORMATION:  
; APPLICANT: PICHUANES, Sergio  
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; FILE REFERENCE: PPI7955.002 / 2301-17955  
; CURRENT APPLICATION NUMBER: US/10/272,459  
; CURRENT FILING DATE: 2002-10-15  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 41  
; LENGTH: 980  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein  
; OTHER INFORMATION: of 115.5 kDa  
US-10-272-459-41

Query Match 100.0%; Score 101; DB 14; Length 980;  
Best Local Similarity 100.0%; Pred. No. 7.4e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VLPPPRKMKGLFSQAKISLF 20  
|||  
DB 823 VLPPPRKMKGLFSQAKISLF 842  
|||

RESULT 4  
US-09-929-955-12  
; Sequence 12, Application US/09929955  
; Patent No. US20020136740A1  
; GENERAL INFORMATION:  
; APPLICANT: Matti Sallberg  
; APPLICANT: Catharina Hultgren  
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND  
; TITLE OF INVENTION: METHODS OF USE THEREOF

FILE REFERENCE: TRIPED.23AUS2  
CURRENT APPLICATION NUMBER: US/09/929,955  
CURRENT FILING DATE: 2001-08-15  
PRIOR APPLICATION NUMBER: 09/705,547  
PRIOR FILING DATE: 2000-11-03  
PRIOR APPLICATION NUMBER: 60/229,175  
PRIOR FILING DATE: 2000-08-29  
PRIOR APPLICATION NUMBER: 60/225,767  
PRIOR FILING DATE: 2000-08-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Hepatitis A virus sequence  
US-09-929-955-12

Query Match 100.0%; Score 101; DB 9; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
DB 823 VLPPPRKMKGLFSQAKISLF 842

#### RESULT 5

US-10-104-966-12  
Sequence 12, Application US/10104966  
Publication No. US20020155124A1  
GENERAL INFORMATION:  
APPLICANT: Catharina Hultgren  
APPLICANT: EMERSON, SUZANNE U  
TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND  
TITLE OF INVENTION: METHODS OF USE THEREOF  
FILE REFERENCE: TRIPED.23AUSC1  
CURRENT APPLICATION NUMBER: US/10/104,966  
CURRENT FILING DATE: 2002-03-22  
PRIOR APPLICATION NUMBER: 09/705,547  
PRIOR FILING DATE: 2000-11-03  
PRIOR APPLICATION NUMBER: 60/229,175  
PRIOR FILING DATE: 2000-08-29  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Hepatitis A virus sequence  
US-10-104-966-12

Query Match 100.0%; Score 101; DB 13; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
DB 823 VLPPPRKMKGLFSQAKISLF 842

#### RESULT 6

US-10-135-988-2  
Sequence 2, Application US/10135988  
Publication No. US20020176869A1  
GENERAL INFORMATION:  
APPLICANT: FUNKHOUSER, ANN W  
APPLICANT: EMERSON, SUZANNE U  
APPLICANT: PURCELL, ROBERT H  
APPLICANT: D'HONDT, ERIC  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES

FILE REFERENCE: 20264262US3  
CURRENT APPLICATION NUMBER: US/10/135,988  
CURRENT FILING DATE: 2002-04-29  
PRIOR APPLICATION NUMBER: 07/947,338  
PRIOR FILING DATE: 1992-09-18  
PRIOR APPLICATION NUMBER: 08/397,232  
PRIOR FILING DATE: 1995-03-10  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: WILD-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175  
US-10-135-988-2

Query Match 100.0%; Score 101; DB 13; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
DB 823 VLPPPRKMKGLFSQAKISLF 842

#### RESULT 7

US-10-135-988-4  
Sequence 4, Application US/10135988  
Publication No. US20020176869A1  
GENERAL INFORMATION:  
APPLICANT: FUNKHOUSER, ANN W  
APPLICANT: EMERSON, SUZANNE U  
APPLICANT: PURCELL, ROBERT H  
APPLICANT: D'HONDT, ERIC  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
FILE REFERENCE: 20264262US3  
CURRENT APPLICATION NUMBER: US/10/135,988  
CURRENT FILING DATE: 2002-04-29  
PRIOR APPLICATION NUMBER: 07/947,338  
PRIOR FILING DATE: 1992-09-18  
PRIOR APPLICATION NUMBER: 08/397,232  
PRIOR FILING DATE: 1995-03-10  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 2227  
TYPE: PRT  
ORGANISM: Attenuated HAV (Pass 35), strain HM-175  
US-10-135-988-4

Query Match 100.0%; Score 101; DB 13; Length 2227;  
Best Local Similarity 100.0%; Pred. No. 1.7e-06;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKISLF 20  
DB 823 VLPPPRKMKGLFSQAKISLF 842

#### RESULT 8

US-10-135-988-6  
Sequence 6, Application US/10135988  
Publication No. US20020176869A1  
GENERAL INFORMATION:  
APPLICANT: FUNKHOUSER, ANN W  
APPLICANT: EMERSON, SUZANNE U  
APPLICANT: PURCELL, ROBERT H  
APPLICANT: D'HONDT, ERIC  
TITLE OF INVENTION: HEPATITIS A VIRUS VACCINES  
FILE REFERENCE: 20264262US3  
CURRENT APPLICATION NUMBER: US/10/135,988  
CURRENT FILING DATE: 2002-04-29  
PRIOR APPLICATION NUMBER: 07/947,338  
PRIOR FILING DATE: 1992-09-18

; PRIOR APPLICATION NUMBER: US/09/171.432A
; FILING DATE: 23-NOV-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Greene, Jamie L.
; REGISTRATION NUMBER: 32,467
; REFERENCE/DOCKET NUMBER: 03063-0231US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 949-2400
; TELEFAX: (404) 949-2499
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..25
; OTHER INFORMATION: /label= YK-1665
US-09-171-432A-47
Query Match 64.4%; Score 65; DB 10; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.0055;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VLPPPRKMKGLF 12
DB 14 VLPPPRKMKGLF 25
RESULT 11
US-09-854-886-2
; Sequence 2, Application US/09854886
; Patent No. US20020072080A1
; GENERAL INFORMATION:
; APPLICANT: Yoshikubo, Takashi
; APPLICANT: Hasegawa, Masami
; TITLE OF INVENTION: Immunological Material and Methods for Detecting
; TITLE OF INVENTION: Dihydropyrimidine Dehydrogenase
; FILE REFERENCE: 100554-32887
; CURRENT APPLICATION NUMBER: US/09/854,886
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 09/138,103
; PRIOR FILING DATE: 1998-08-21
; PRIOR APPLICATION NUMBER: 97114630.3
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1025
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-886-2
Query Match 52.0%; Score 52.5; DB 9; Length 1025;
Best Local Similarity 63.2%; Pred. No. 22;
Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;
QY 2 LPPPRKMKGLFSQAKISLF 20
DB 176 LPPPRKMKSEAYS-AKIALF 193
RESULT 12
US-10-424-599-242191
; Sequence 242191, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:

; PRIOR APPLICATION NUMBER: 08/397,232
; PRIOR FILING DATE: 1995-03-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 6
; LENGTH: 2227
; TYPE: PRT
; ORGANISM: Attenuated (4380) HAV, strain HM-175
US-10-135-988-6
Query Match 100.0%; Score 101; DB 13; Length 2227;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VLPPPRKMKGLFSQAKISLF 20
DB 823 VLPPPRKMKGLFSQAKISLF 842
RESULT 9
US-10-272-459-40
; Sequence 40, Application US/10272459
; Publication No. US20030124517A1
; GENERAL INFORMATION:
; APPLICANT: PICHDANTES, Sergio
; TITLE OF INVENTION: HEPATITIS A VIRUS NUCLEOTIDE SEQUENCES, RECOMBINANT
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: P17955.002 / 2301-17955
; CURRENT APPLICATION NUMBER: US/10/272,459
; CURRENT FILING DATE: 2002-10-15
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40
; LENGTH: 836
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: recombinant protein
; OTHER INFORMATION: of 94 kDa
US-10-272-459-40
Query Match 69.3%; Score 70; DB 14; Length 836;
Best Local Similarity 92.9%; Pred. No. 0.037;
Matches 13; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VLPPPRKMKGLFSQ 14
DB 823 VLPPPRKMKGLFSQ 836
RESULT 10
US-09-171-432A-47
; Sequence 47, Application US/09171432A
; Publication No. US20030187184A1
; GENERAL INFORMATION:
; APPLICANT: Fields, Howard A.
; APPLICANT: Knudsyakov, Yuri E.
; TITLE OF INVENTION: Antigenically Reactive Regions of the
; TITLE OF INVENTION: Hepatitis A Virus Polypeptide
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kilpatrick Stockton LLP
; STREET: 3424 Peachtree Road, N.E.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30326
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 242191
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_60726C.1.pep
US-10-424-599-242191

Query Match          50.5%; Score 51; DB 12; Length 84;
Best Local Similarity 66.7%; Pred. No. 2.8;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 PPPRKMKGFLFSQAKI 17
    |||:|||||
Db 59 PPPKKKGGFSRPKI 73

RESULT 13
US-10-424-599-264591
; Sequence 264591, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 264591
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_80946C.1.pep
US-10-424-599-264591

Query Match          47.5%; Score 48; DB 12; Length 187;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 PPPRKMKGFLFSQAKISLF 20
    |||:|||||
Db 122 PPPRGFVGLFSRPTGFF 139

RESULT 14
US-10-320-797-3221
; Sequence 3221, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Broshtkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
```

```
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3221
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Cryptococcus neoformans
US-10-320-797-3221

Query Match          47.5%; Score 48; DB 15; Length 524;
Best Local Similarity 43.8%; Pred. No. 55;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 LPPPRKMKGFLFSQAKI 17
    |||:|||||
Db 5 LPPPKRQKSAYSQSQM 20

RESULT 15
US-10-424-599-280225
; Sequence 280225, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 280225
; LENGTH: 650
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_95066C.1.pep
US-10-424-599-280225

Query Match          46.5%; Score 47; DB 12; Length 650;
Best Local Similarity 62.5%; Pred. No. 97;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 LPPPRKMKGFLFSQAKI 17
    |||:|||||
Db 85 LPDPFKKGRFTQKKI 100

RESULT 16
US-10-424-599-196137
; Sequence 196137, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196137
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19137C.1.pep
US-10-424-599-196137

Query Match          45.5%; Score 46; DB 12; Length 73;
```



```

Best Local Similarity 50.0%; Pred. No. 14;
Matches 9; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 VLPPPRKMKGLFSQAKIS 18
Db 12 IAPPPKXKKGESPLAKIT 29
: |||: |||:
: |||: |||:

RESULT 17
US-10-424-598-265499
; Sequence 265499, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 265499
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(158)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_81766C.1.pep
US-10-424-599-265499

Query Match 45.5%; Score 46; DB 12; Length 158;
Best Local Similarity 52.9%; Pred. No. 32;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQAKISL 19
Db 92 PPKKKKGGLKEPKVL 108
|||: |||: |||:
|||: |||: |||:

RESULT 18
US-10-094-749-2193
; Sequence 2193, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KENJI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24

```

; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_73833C.1.pap  
US-10-424-599-256711

Query Match 44.6%; Score 45; DB 12; Length 2834;  
Best Local Similarity 47.4%; Pred. No. 9.2e+02;  
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 VLPPPRKMGKGLFSQAKISL 19  
DB 1109 LMERVRKLGKGLASRKASI 1127

RESULT 21  
US-10-424-599-200960  
; Sequence 200960, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 200960  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(151)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_23493C.1.pap  
US-10-424-599-200960

Query Match 43.6%; Score 44; DB 12; Length 151;  
Best Local Similarity 52.6%; Pred. No. 6;  
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 VLPPPRKMGKGLFSQAKISL 19  
DB 76 LPPSTPTSTPLFSQKINL 94

RESULT 22  
US-10-424-599-198250  
; Sequence 198250, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 198250  
; LENGTH: 178  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_21044C.1.pap  
US-10-424-599-198250

Query Match 43.6%; Score 44; DB 12; Length 178;  
Best Local Similarity 45.0%; Pred. No. 73;  
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 VLPPPRKMGKGLFSQAKISLF 20  
DB 158 VRPPTRLKTKTYKATPNLF 177

RESULT 23  
US-10-087-684-76  
; Sequence 76, Application US/10087684  
; Publication No. US20040029116A1  
; GENERAL INFORMATION:  
; APPLICANT: Edinger, Shlomit R.  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Stone, David J.  
; APPLICANT: Grosse, William M.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: Rieger, Daniel K.  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Casman, Stacie, J.  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Li, Li  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Zerhusen, Bryan D.  
; APPLICANT: Malyankar, Uriel M.  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Gangoli, Esha A.  
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 21402-214 CIP  
; CURRENT APPLICATION NUMBER: US/10/087,684  
; CURRENT FILING DATE: 2003-03-10  
; PRIOR APPLICATION NUMBER: 60/253,834  
; PRIOR FILING DATE: 2000-11-29  
; PRIOR APPLICATION NUMBER: 60/250,926  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 60/264,180  
; PRIOR FILING DATE: 2001-01-25  
; PRIOR APPLICATION NUMBER: 60/274,194  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/313,656  
; PRIOR FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/327,456  
; PRIOR FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 220  
; SOFTWARE: Curaseqlist version 0.1  
; SEQ ID NO 76  
; LENGTH: 408  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (274)..(274)  
; OTHER INFORMATION: Wherein Xaa may be any naturally occurring amino acid  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (341)..(341)  
; OTHER INFORMATION: Wherein Xaa may be any naturally occurring amino acid  
US-10-087-684-76

Query Match 43.6%; Score 44; DB 12; Length 408;  
Best Local Similarity 50.0%; Pred. No. 1.7e+02;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

```
QY      3 PPRKMKGLFSQAKISLF 20
DB      186 PDPKMTQLSQAKLAVF 203

RESULT 24
US-10-218-779-76
; Sequence 76, Application US/10218779
; Publication No. US20040029222A1
; GENERAL INFORMATION:
; APPLICANT: Edinger, Shlomit
; APPLICANT: MacDougall, John
; APPLICANT: Millet, Isabelle
; APPLICANT: Ellerman, Karen
; APPLICANT: Stone, David
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William
; APPLICANT: Alsobrook II, John
; APPLICANT: Lepley, Denise
; APPLICANT: Rieger, Daniel
; APPLICANT: Burgess, Catherine
; APPLICANT: Casman, Stacie
; APPLICANT: Spytek, Kimberly
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Mishra, Vishnu
; APPLICANT: Patturajan, Meera
; APPLICANT: Shency, Suresh
; APPLICANT: Rastelli, Luca
; APPLICANT: Tchernev, Velizar
; APPLICANT: Vernet, Corine
; APPLICANT: Zehusen, Bryan
; APPLICANT: Malyankar, Uriel
; APPLICANT: Guo, Xiaojia
; APPLICANT: Miller, Charles
; APPLICANT: Gangolli, Esha
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-214
; CURRENT APPLICATION NUMBER: US/10/218,779
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 60/253,834
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/250,-926
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/264,180
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/313,656
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/327,456
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 76
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-218-779-76

Query Match      43.6%; Score 44; DB 12; Length 408;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY      3 PPRKMKGLFSQAKISLF 20
DB      186 PDPKMTQLSQAKLAVF 203

RESULT 25
US-10-425-114-54364
; Sequence 34364, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 54364
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73030H10_FLI.pgp
US-10-425-114-54364

Query Match      43.6%; Score 44; DB 12; Length 457;
Best Local Similarity 77.8%; Pred. No. 1.9e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      3 PPRKMKGL 11
DB      207 PPRKMKGL 215

RESULT 26
US-10-425-114-43063
; Sequence 43063, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 43063
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB143-061-G7_FLI.pgp
US-10-425-114-43063

Query Match      43.6%; Score 44; DB 12; Length 490;
Best Local Similarity 77.8%; Pred. No. 2.1e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      3 PPRKMKGL 11
DB      240 PPRKMKGL 248

RESULT 27
US-10-104-047-3665
; Sequence 3665, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
```

```
; CURRENT FILING DATE: 2002-03-25
; PRIOR FILING DATE:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3665
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3665

Query Match      43.6%; Score 44; DB 15; Length 522;
Best Local Similarity 47.1%; Pred. No. 2.2e+02;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQAKISL 19
Db 400 PPRKMKGLFSQAKISL 416

RESULT 28
US-10-176-306-2
; Sequence 2, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Glucksmann, Maria Alexandra
; APPLICANT: Bandaru, Rajasekhar
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
; FILE REFERENCE: 10448-195001
; CURRENT APPLICATION NUMBER: US/10/176,306
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: 10/001,137
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: PCT/US01/45291
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/248,362
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,331
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,365
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/250,077
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,327
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,176
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 10/023,617
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: PCT/US01/49416
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/256,249
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/256,405
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 10/083,248
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/US01/46717
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/242,324
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,518
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/241,989
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 673
; TYPE: PRT

; ORGANISM: Homo sapiens
US-10-176-306-2

Query Match      43.6%; Score 44; DB 14; Length 673;
Best Local Similarity 43.8%; Pred. No. 2.9e+02;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPRKMKGLFSQAKIS 18
Db 507 PPRKMKGLFSQAKIS 522

RESULT 29
US-10-424-599-212719
; Sequence 212719, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 212719
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_34110C.1.pep
US-10-424-599-212719

Query Match      43.1%; Score 43.5; DB 12; Length 62;
Best Local Similarity 43.5%; Pred. No. 29;
Matches 10; Conservative 4; Mismatches 6; Indels 3; Gaps 1;

QY 1 VLPPPKMKG---LFSQAKISLF 20
Db 32 LLPPPKKSQGNYPFSLNKAVFF 54

RESULT 30
US-09-933-767-855
; Sequence 855, Application US/09933767
; Publication No. US20030181692A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007P2
; CURRENT APPLICATION NUMBER: US/09/933,767
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: PCT/US01/05614
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,836
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/193,170
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 09/205,258
; PRIOR FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: PCT/US98/11422
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/048,885
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/049,375
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/048,881
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/048,880
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/048,896
```

;  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/049,020  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,876  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,895  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,884  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,894  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,971  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,964  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,882  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,899  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,893  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,900  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,901  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,892  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,915  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/049,019  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,970  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,972  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,916  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/049,373  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,875  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/049,374  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,974  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,883  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,897  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,898  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,962  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,963  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,877  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/048,878  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/068,054  
; PRIOR FILING DATE: 1997-12-18  
; PRIOR APPLICATION NUMBER: 60/068,064  
; PRIOR FILING DATE: 1997-12-18  
; PRIOR APPLICATION NUMBER: 60/068,053  
; PRIOR FILING DATE: 1997-12-18  
; PRIOR APPLICATION NUMBER: 60/070,923  
; PRIOR FILING DATE: 1997-12-18  
; PRIOR APPLICATION NUMBER: 60/073,160  
; PRIOR FILING DATE: 1998-01-30

;  
; PRIOR APPLICATION NUMBER: 60/073,159  
; PRIOR FILING DATE: 1998-01-30  
; PRIOR APPLICATION NUMBER: 60/073,165  
; PRIOR FILING DATE: 1998-01-30  
; PRIOR APPLICATION NUMBER: 60/073,164  
; PRIOR FILING DATE: 1998-01-30  
; PRIOR APPLICATION NUMBER: 60/085,925  
; PRIOR FILING DATE: 1998-05-18  
; PRIOR APPLICATION NUMBER: 60/085,921  
; PRIOR FILING DATE: 1998-05-18  
; PRIOR APPLICATION NUMBER: 60/085,923  
; PRIOR FILING DATE: 1998-05-18  
; PRIOR APPLICATION NUMBER: 60/085,922  
; PRIOR FILING DATE: 1998-05-18  
; PRIOR APPLICATION NUMBER: 60/092,921  
; PRIOR FILING DATE: 1998-07-15  
; PRIOR APPLICATION NUMBER: 60/094,657  
; PRIOR FILING DATE: 1998-07-30  
; NUMBER OF SEQ ID NOS: 1245  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 855  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-933-767-855

Query Match 43.1%; Score 43.5; DB 10; Length 97;  
Best Local Similarity 47.4%; Pred. No. 46;  
Matches 9; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

QY 2 LPPRRKMKGLFSQAKISLF 20  
:|||||:|:|:|:  
Db 52 VPPFLKMPGVLA-IELSIF 69

Search completed: May 11, 2004, 13:44:51  
Job time : 35.25 secs